PRIVACY-AWARE SURVEILLANCE SYSTEM DESIGN

MUKESH KUMAR SAINI

NATIONAL UNIVERSITY OF SINGAPORE

2012

PRIVACY-AWARE SURVEILLANCE SYSTEM DESIGN

MUKESH KUMAR SAINI

(M.Tech), CEDT

IISc Bangalore, India

A THESIS SUBMITTED

FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF COMPUTER SCIENCE
SCHOOL OF COMPUTING

NATIONAL UNIVERSITY OF SINGAPORE

2012

To my family & my beloved Guddu.

Acknowledgment

My course of PhD has been a great learning experience where many people taught me the importance of perseverance and focus in research work. I would like to thank my PhD supervisor, Dr. Mohan Kankanhalli, for his continuous support and encouragement. He has been patient with my many mistakes, and provided me appropriate guidance to learn from those mistakes and overcome them. I take this opportunity to express my sincere gratitude to all my collaborators who have been a source of great motivation and learning for me. I want to thank Dr. Pradeep Atrey, Dr. Sharad Mehrotra, and Dr. Ramesh Jain for giving me the opportunity to work with them and at the same time providing insights in the art of doing research. Their involvement has been important for me in cultivating interest on various topics of multimedia systems and analytics. I have learnt a lot from my interactions with my supervisor and my collaborators, especially in the way to conduct research. I also express my deepest gratitude to members of my thesis committee, Dr. Roger Zimmermann and Dr. Wei Tsang Ooi, for their efforts and valuable input at different stages of my PhD. Finishing my research work would not be possible without the support of family and friends. I want to thank my parents, brothers, and sister for their unconditional moral support at all times during my graduate student life. All my friends have been extremely generous with their encouragement and motivation. I enjoyed my numerous discussions with Li Zhonghua, Gan Tian, Wang Xiangyu, Dwarikanath Mahapatra, and Harish Katti on various topics of my research. Finally, I would like to thank all the anonymous reviewers whose comments helped me to improve my papers and present my research better to a large audience.

Abstract

Video surveillance is a very effective means of monitoring activities over a large area with cameras as extended eyes. However, this additional security comes at the cost of privacy loss of the citizens not involved in any illicit activities. The traditional privacy protection methods only consider facial cues for identity leakage and privacy loss. Because an adversary can use prior knowledge to infer the identities even in the absence of the facial information, we propose a privacy-aware surveillance framework in which we identify the implicit channels that cause identity leakage, quantify privacy loss through non-facial information, and propose solutions to block these channels for near zero privacy loss with minimal utility loss. Privacy loss is modeled as an adversary's ability to correlate sensitive information to the identity of the individuals in the video. Anonymity based approach is used to consolidate the identity leakage through explicit channels of bodily cues such as facial information; and other implicit channels that exist due to *what*, *when*, and *where* information. The proposed privacy model is applied to two important applications of surveillance video data publication and CCTV monitoring. Through experiments it is found that current privacy protection methods include high risk of privacy loss while the proposed framework provides more robust privacy loss measures and better tradeoff of security and privacy.

Contents

List of Symbols

Γ	Privacy Loss
I	Identity Leakage
$\mathcal{I}$	Identity Leakage Vector
I_{im}	Implicit Identity Leakage
I_{ex}	Explicit Identity Leakage
I_{wo}	Identity Leakage due to *who* evidence
I_{wt}	Identity Leakage due to *what* evidence
I_{wn}	Identity Leakage due to *when* evidence
I_{wr}	Identity Leakage due to *where* evidence
I_{wo}	Identity Leakage due to *who* evidence
G_{wt}	Association group size for given *what* evidence
I_{wtwr}	Association group size for given *what* and *where* evidences

I_{wtwn}	Association group size for given *what* and *when* evidences
I_{wtwnwr}	Association group size for given *what*, *when* and *where* evidences
$\mathcal{S}$	Sensitivity Vector
S	Sensitivity Matrix
Ψ	Sensitivity Index
$\mathcal{W}$	Priority Vector
τ	Video Analysis Task
V	Original Video Data
V'	Transformed Video Data
U	Utility Loss
E	Energy function
$\mathcal{F}$	Data Transformation Function
η	Importance Factor
α	Task Weight
f	Video frames

b	Degree of blurring
q	Degree of quantization
$\mathbf{G}$	Set of association groups
$\mathcal{G}$	Association groups
$\mathcal{P}$	Proposition of a logical statements
$\mathcal{C}$	Conclusion of a logical statements
κ	Anonymity
$\mathcal{R}$	Set of user ratings
Λ	Absence Privacy
χ	Transition Matrix
$\mathcal{Z}$	State space for Markov Chain
$\mathcal{H}$	People count function
TFG	Target Flow Graph
E_{pc}	Equalization function

List of Figures

List of Tables

Chapter 1

Introduction

Because an adversary can use prior knowledge to infer the identities of individuals in the video even in the absence of facial information, we develop a privacy-aware surveillance framework in which we identify the implicit channels of identity leakage, quantify the privacy loss through non-facial information, and propose solution to block these channels for near zero privacy loss with minimal utility loss. The proposed privacy loss model considers facial as well as non-facial information and is able to consolidate the identity leakage through multiple events and multiple cameras. Moreover, any privacy preserving method usually affects the utility of the data; therefore, the choice of data transformation is paramount to ensure an acceptable tradeoff between the privacy and the utility. We propose utility models and privacy preservation framework for the applications of surveillance and data publication.

1.1 Motivation

Security concerns are increasing rapidly at both public and private places. Recent terrorist attacks have intensified the security demands in the society. The violation of law and order is unfortunately common in most of the major cities in the world. The quick rise in such illegal activities and increased number of offenders have forced the governments to make personal and asset security a high priority task in their policies. To combat these security concerns, it is needed to monitor all public places, commercial venues, and military areas. Therefore, multimedia surveillance has a wide spectrum of promising applications, for example traffic surveillance in cities, detection of military targets, and a security defender for communities and

important buildings [Rat10]. A large number of cameras are being installed to increase the coverage area of the surveillance operators. The growing number of surveillance cameras is causing privacy loss of people not involved in any wrong doings [Cav07].

Privacy is a big concern in current video surveillance systems. Due to privacy concerns, many strategic places remain unmonitored, leading to security threats. With respect to surveillance video, there are mainly two places where privacy loss could occur: when security personnel are watching the video currently being captured by the cameras, and when the recorded video is disseminated for forensics and other research purposes. For both the cases, the first step is to analyze the characteristics of the video which cause privacy loss (privacy modeling) and the second step is to modify the video data (data transformation) to preserve the privacy. To accomplish these steps, we need to model and quantify privacy loss and utility loss of the video data.

In the past, the problem of privacy preservation in video has been addressed mainly by surveillance researchers. Specifically, computer vision techniques are used extensively to first detect the faces in the images and then obfuscate them [NSM05]; however, other implicit inference channels through which an individual's identity can be learned have not been considered. An adversary can observe the behavior, look at the places visited, and combine that with the temporal information to infer the identity of the person in the video. Consider a school in which Prof Pradeep and Prof Ramesh are the only staff members who eat in the vegetarian canteen and Prof Ramesh is a visiting faculty member who only comes in the afternoon. With this knowledge, an adversary can observe that person X has been spotted at the staff club as well as the vegetarian canteen and infer that person X is either Prof Pradeep or Prof Ramesh, even without having the facial information. In addition, the adversary also knows that current time is morning, s/he can further infer that X is indeed Prof Pradeep.

1.2 Background

The main goal of surveillance is to ensure safety and security of the citizens. However, it is impossible for security personnel like police forces to manually monitor all the places physically. Therefore, multimedia sensors are employed as an aid to the surveillance operator as shown in Figure 1.1. Particularly, current surveillance systems use large number of cameras [Lib07]

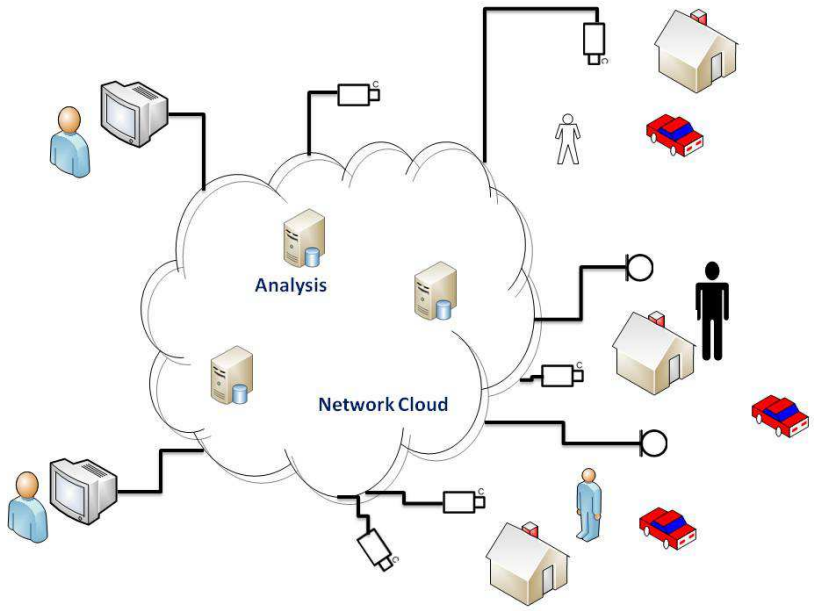


Figure 1.1: A typical video surveillance system. Video cameras and microphones capture the events and activities of the environment.

(Figure 1.2) to assess the situation and reduce security threats. However, this safety comes at the cost of privacy of the individuals not involved in any illicit activities. Privacy concerns prohibit us from keeping cameras at many critical places which need to be monitored. Still, a large number of cameras are being placed to increase the coverage area. The huge amounts of video recorded by surveillance cameras is generally discarded due to privacy concerns. Video is capable of recording and preserving enormous amount of information that can be used in many applications ranging from forensics to ethnography and other behavioral studies. Therefore, in this thesis we analyze the privacy loss that might occur due to public access of the surveillance video.

Protecting privacy of the individuals is important. Many countries have identified privacy as a fundamental right and have attempted to make it a law [Ass48]. The oldest known legislation on privacy is England's 1361 Justices of the Peace Act against eavesdroppers and stalkers [BS03]. In 1890, US Supreme Court Justice Louis Brandeis recognized privacy as "the right to be left alone" and declared it to be fundamental right of democracy [BZK+90]. Chesterman [Che11] explains the need for collection of surveillance data and proposes the civilians to accept this loss of privacy as the cost of security. The Global Internet Liberty Campaign [BD99] has done an survey extensive to divide privacy in broadly four categories:

Figure 1.2: Multiple cameras installed at a single site.

- **Information Privacy**: personal data such as credit card information and medical records;
- **Bodily privacy**: it concerns bodily attributes e.g. cavity and internal injuries;
- **Privacy of communications**: conversations via mail, telephones, email and other forms of communication;
- **Territorial privacy**: location information such as places visited by an individual.

The surveillance video generally causes loss of "Bodily privacy" and "Territorial privacy". However, the sense of privacy is a subjective affair and it may depend on the individual's habits, preferences, and moral views [Lan01]. We extend these categories to include companion, activity, and appearance as potentially sensitive information that can cause privacy loss.

1.3 Issues In Privacy-Aware Use of Surveillance Video

The main contributing factors of the privacy loss are the identity leakage of the individuals and the presence of the sensitive information. Below are the important terms and definitions that will be used in thesis with respect to the identity leakage.

Definition 1 ***Explicit Channels****: These are the bodily identity leakage cues that are used to identify a person. They mainly include facial and appearance information.*

Definition 2 ***Implicit Channels****: These are the contextual cues used to identify a person. They mainly include what, when, and where information.*

Definition 3 ***Identity Leakage****: The certainty with which an adversary can identify an individual in the video. It is equivalent to inverse of the anonymity of individuals.*

Definition 4 ***Explicit Identity Leakage****: The identity leakage due to explicit channels.*

Definition 5 ***Implicit Identity Leakage****: The identity leakage due to implicit channels.*

We recognize that the privacy loss is association of the identity with certain information in the video that might be sensitive to the individuals. The knowledge of the identity of a person is modeled as the identity leakage while the presence of sensitive information is denoted as the sensitivity index:

Definition 6 ***Sensitivity Index****: This is a measure of sensitive information in the video for which an individual feels privacy violation would occur if made available to public.*

Following are the important issues need to be considered for privacy-aware surveillance system and privacy-aware publication of surveillance video data.

1.3.1 What causes privacy violation?

These issues are concerned with the robust privacy modeling which is necessarily the first step of any privacy protection method.

1. **Sensitive Information Vs Identity** In early days of video conferencing, it was understood that video always contains the sensitive information about the individuals and it was transformed (e.g. blurred) such that users could know the identity, but could not access the the sensitive information. In the current surveillance scenarios, the identity of a person is modeled as privacy loss. Many methods of hiding the identity have been proposed (e.g. face or blob obfuscation). We argue that both the sensitive information and the identity should be considered to measure the privacy loss. The challenge here is how to quantify these properties and combine to obtain overall privacy loss.

2. **Implicit Channels Vs Explicit Channels** In the past, the facial information is considered as main source of the identity leakage. While blocking the facial information is necessary for preserving the identity, it is not sufficient. Identity could also be inferred

through non-facial information like time, place, events, and activities. It is important to quantify the identity leakage through these implicit channels in order to provide robust privacy measures.

3. **Single Camera/Multiple Camera** The adversary (a surveillance operator or a person using published video) might have access to video from multiple cameras. Multiple cameras can provide additional information through correlated events and activities. If the adversary has the knowledge of usual event and activity patterns at the surveillance premise, access to multiple cameras might further increase the identity leakage. How to quantify this additional identity leakage and combine with the identity leakage from single cameras is a research challenge.

1.3.2 How to transform data to reduce privacy loss?

Once we get the tool to measure the privacy loss, we need to transform the data such that the privacy is preserved with minimal compromise in the utility. Further, the surveillance data is gigantic in size; therefore, the process of data transformation needs to be automated enough to avoid the scalability problem. This requires the following issues to be considered:

1. **Utility Measurement** Earlier works on privacy-aware application of video data have mainly focused on the robustness of the privacy protection methods. However, there are many transformation methods to achieve similar levels of privacy preservation. To decide which method provides best tradeoff between privacy and intended application of the data, we need to quantify the utility of the original and transformed video data.

2. **Transformation Method** The video data can be transformed in multiple ways e.g. pixelization, quantization, and blurring. Many a times, using a combination of transformation functions may give better tradeoff between the utility and the privacy. The question here is how to choose the transformation function and in what order they should be applied on the video data.

3. **Selective Obfuscation Vs Global Operations** In order to hide the privacy information, the video data needs to be transformed. There can be two approaches of transforming the data. In the first approach the privacy regions are determined with help of computer

vision detectors and obfuscated. The biggest problem with this approach is that detectors may fail, providing a non-robust privacy protection. The second approach is to globally transform the whole image to hide the privacy information. This approach is very pessimistic and results in huge utility loss. A combination of both the approaches should be used to obtain optimal tradeoff between the privacy and the utility of the video data.

1.4 Thesis Contributions

The main contributions of thesis are on building models to quantify the privacy loss and utility loss; and their application in privacy-aware surveillance and data publication as follows:

1. The past works have considered only explicit identity leakage (mainly facial information), but they have not taken into account the implicit channels of *what, when* and *where*. To the best of our knowledge, this is the first attempt to model the privacy loss as a continuous variable considering both explicit and implicit channels.

2. Most of the earlier works have modeled the privacy loss as the identity leakage alone or presence of the sensitive information alone. However, the privacy loss is a function of both of these quantities. In this thesis we quantify the identity leakage and the sensitivity index, and propose model to combine these quantity and calculate overall privacy loss.

3. We model the utility loss for the application of data publication and propose a hybrid data transformation method (using a combination of quantization and blurring). This provides an opportunity of publishing surveillance video data which can be very useful for testing vision algorithms, video ethnography, data mining, and policy making.

4. In the traditional surveillance system, the CCTV operator has prior context knowledge of the surveillance site and its habitants, which makes it difficult to block the implicit identity leakage channels. We propose an anonymous surveillance framework that advocates decoupling the contextual knowledge from the video. The experiments show that the proposed framework is effective in blocking the identity leakage channels and provides better sense of privacy to the individuals.

5. The surveillance task is target (people, vehicle, etc.) centric and the amount of human

attention depends on the number of the targets in the camera view. We model this workload as a Markov chain and propose a dynamic workload assignment method that equalizes the number of targets monitored by each operator by dynamically changing the camera-to-operator assignment.

1.5 Thesis Organization

The thesis is organized as follows. In Chapter 2 a review of the related works is provided. We review the earlier works from privacy modeling and data transformation perspectives. We also provide a review of the existing video datasets and their limitations. Two tables have been provided to precisely compare the proposed work with the earlier works. In Chapter 3 we provide a privacy model for video from a single camera. The model combines the identity leakage from the implicit and the explicit channels. The model is applied in the scenarios of privacy-aware video data publication. Extensive experiments are provided to demonstrate the method.

An enhanced model of the privacy loss for multi-camera scenario is proposed in Chapter 4. In this model, we use an event based framework to measure the identity leakage from multiple cameras. The findings from the privacy models are applied to traditional surveillance systems in Chapter 5. It is found that in the current surveillance systems it is very difficult to hide the identity information from the CCTV operator. Consequently, an anonymous surveillance framework is proposed that decouples the CCTV operator's contextual knowledge from the video data; and ensures enhanced privacy protection. Chapter 6 provides a summary of thesis, conclusions, and it ends with the future research challenges that need to be solved in order to provide robust privacy loss measures.

Chapter 2

Related Work

We review the privacy works from two perspectives: privacy modeling and data transformation. In the privacy modeling, we describe different methods used to measure the privacy loss and compare them with our proposed model. Next we discuss privacy protection methods employed in various surveillance systems to understand their limitations. Finally, the need to publish real surveillance data is emphasized by analyzing the existing datasets. We have also provided a brief review of the privacy works in statistical data publication to form a background for anonymity based privacy modeling.

2.1 Privacy Modeling

As the main focus of thesis is to design a privacy-aware video surveillance system, the first step is to understand what characteristics of a video cause privacy loss. There has been only little work specifically on privacy modeling. However, all the works on privacy-aware use of multimedia assume some model of privacy loss in their framework. In this section we analyze these works and discuss their robustness and adequacy for surveillance video. We have divided the works into two broad categories. In the first set of works, it is assumed that the identity is known through other means and the semantic information of the video causes privacy loss. The other set of works assume that the identity leakage itself is equivalent to the privacy loss.

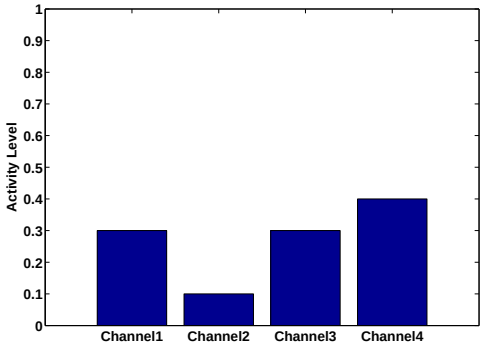


Figure 2.1: Activity bars for four channels.

2.1.1 Sensitive Information as Privacy Loss

Privacy modeling in terms of the sensitive information is prevalent in the fields of office video conferencing [DB92, FKRR93, TIR94] and pervasive computing [BS03, BA07] scenarios where the identity of individuals is generally known to the adversary due to the user centric nature of the applications. The goal of the researchers is to perform the intended application (i.e. video conferencing and pervasive computing), without exposing the sensitive information. For instance, some users are sensitive to their activity or location information. Hence, in these works the privacy is modeled as the presence of such sensitive information:

$$\Gamma = \begin{cases} 0 & \text{if video does not contain any sensitive information.;} \\ 1 & \text{otherwise.} \end{cases} \tag{2.1}$$

Ackerman et al. [AS95] use iconic representation of humans in place of of actual images. The authors use activity bars to indicate the activity level of the individuals participating in the office conference scenario as shown in the Figure 2.1. Similarly, in NYNEX porthole system [LGS97] [LSG97] the privacy is preserved by just showing the activity change in form of color bars. The activity change is detected by comparing pixel values between consecutive frames. As users may be interested in knowing the activity level over a window of time, the activity bars are displayed for multiple instants (Figure 2.2). The system also provides its users a control to blur the images globally, which are displayed along with the activity bars. In this work also it is assumed that the privacy loss occurs due to presence of the sensitive information and blur helps in removing the details of the image which might be sensitive. Zhao and Stasko [ZS98] filter the video so that the individuals are identified, but other sensitive information like where they are,

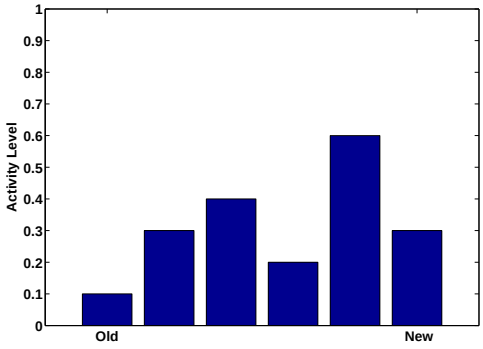


Figure 2.2: Activity bars at six consecutive instants for a given channel.

what they are doing, or who are with them, cannot be detected. To do that, the authors filter the videos using three techniques: pixelization, edge detection, and shadow-view. Still, these works do not provide computational models for the privacy or the utility of the data.

The most common intrusion in pervasive computing environments is users location information and the ability to track it over extended period of time. Attempts have been made to separate the location information from the identities of the users in pervasive computing environments [CAMN+03, AMCK+02]. Beresford and Stajano [BS03] use pseudonyms and mix zones to anonymize the identity and location of the users seeking location aware services. Pseudonyms and anonymous identifiers are assigned to the user by a middleware between the user and the service provider. It is argued in [BS03] that pseudonyms alone are not sufficient for preserving the identity and the it can be inferred by analyzing the time spent by each user at various locations.

Cheng et al. [CZT05] use false data to anonymize the identity. For instance, to receive location aware advertisements on mobile phone and also protect location privacy at the same time, false requests with random sites are also sent with the true request. In this way, the service provider cannot know the users location precisely. The user can choose the correct advertisement to see. Similarly, Bhaskar and Ahamed [BA07] describe obstructive nature of pervasive computing environments. They also discuss the privacy issues which arise due to location and context dependency of pervasive computing. Patrikakis [PKV07] provides a broader review of privacy concerns in pervasive computing. Zhu et al. [ZCZM10] propose to reveal bare minimum amount of user information to the service provider. For example, if a service needs to know the city of residence, it should only be provided with city name rather than

full home address. A user centric privacy protection framework is provided in [BZK+10] where users can define their own priorities for privacy and negotiate with the service providers. In pervasive computing, users can define their privacy policies, which are implemented accordingly by mapping on the data level privacy policies (how to transform the data) [DUM10].

Tansuriyavong and Hanaki [TH01] propose a privacy preserving method for circumstantial videos in which the human body is replaced by silhouette to protect the privacy. The name of the individuals is displayed as text on the silhouette. In this work, the authors implicitly assume that people are sensitive to appearance information. However, in normal scenarios people can also be sensitive to other type of information as well such as time, place, and companion. There are mainly two problems with these works. Firstly they do not tell how to quantify the sensitivity of the video data, and secondly they don't study how the uncertainty in the identity information affects the overall privacy loss.

In contrast to surveillance, pervasive computing is a user centric application where the individual's identity is generally known, while surveillance is an event centric application where many tasks can be performed without knowledge of the identity. We agree that an adversary's end goal is to know the sensitive information about the individuals, still, if the identity is preserved, the sensitive information alone does not cause privacy loss. This is similar to the statistical data publication (e.g. hospitals publishing medical records for research purposes), where they remove the identifiers like name, SSN number, and insurance number before publishing the data. In this way, the sensitive information like the name of disease and age of the patient is exposed but the identities are preserved. It is difficult to hide the identity in pervasive computing and video conferencing, however, as discussed earlier, many surveillance tasks can be done without the identity information. For example, a suspicious activity, fight, intrusion, or stampede can be detected without the identity information. Hence, there is a need to study the combined effect of the identity leakage and the sensitive information on the overall privacy loss.

2.1.2 Identity as Privacy Loss

Measuring the privacy loss in terms of the identity leakage is the most common approach of privacy modeling in video surveillance community. In this approach, it is assumed that if

Figure 2.3: The motion information is superimposed on the static background. Dark colored boxes represent more recent motion.

the adversary can recognize a person in the video, it always leads to privacy loss. With this assumption, they model the privacy loss in terms of the characteristics of the video which reveal the identity of the person. The most common approach of privacy preservation has been to detect and obfuscate the facial regions.

Hudson and Smith [HS96] take the reference image of the static background and superimpose the human motion information in form of dark squares (Figure 2.3). The authors also discuss the removal of the privacy information from the audio data. In that, they remove the intelligible words from the speech and equalize the speech volume. Boyle et al. [BEG00] evaluate the effect of blurring and pixelization on the privacy protection. Through user studies, they found blurring provides better tradeoff between the utility of the data and the privacy loss, however, no models are provided to measure these quantities. Wickramasuriya et al. [WDMV04] define four levels of privacy: original image, blurred silhouette, monotonically colored silhouette, and bounding box as shown in Figure 2.4. They detect the authorized people using motion sensors and RFID tags, who are subsequently masked to protect privacy.

Zhang et al. [ZCC05] detect the human bounding boxes in the images and replace these bounding boxes by the background. The background is estimated using Kalman filtering [KvB90]. The extracted private information is separately encoded and then embedded into the original video as watermark. Cheung et al. [CPN08, CVP+09] use object detection to determine the bounding boxes covering the whole human body and replace these regions with background. The original data of the bounding box is encrypted and sent with the obfuscated data, which can be seen by authorized person's using the encryption key. Most of these works assume a binary model of privacy, where hiding human silhouette, or obfuscating bounding box

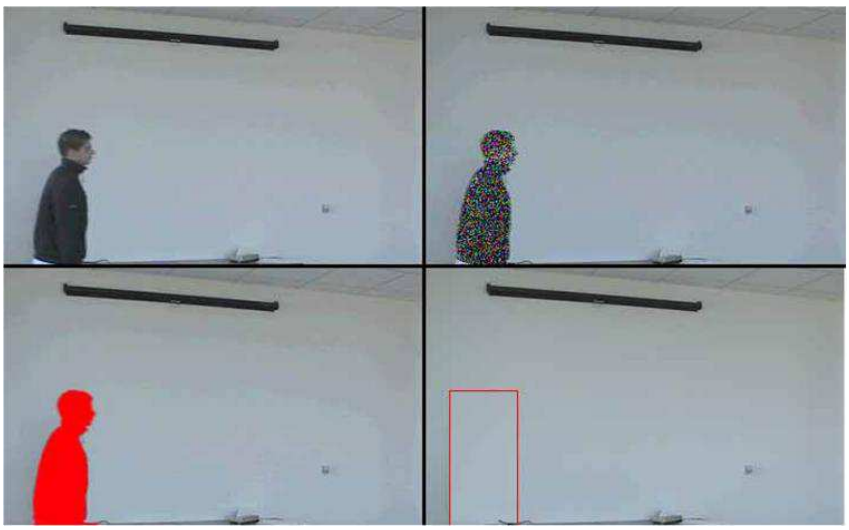


Figure 2.4: Four levels of privacy: original, noisy/blurred, pixel colorized, and bounding box. Image taken from [WDMV04]

covering human is considered as zero privacy loss, otherwise full privacy loss:

$$\Gamma = \begin{cases} 0 & \text{if the human silhouette or bounding box is obfuscated;} \\ 1 & \text{otherwise.} \end{cases} \tag{2.2}$$

Berger [Ber00] describes the privacy model in which the user specifies the skin tone of the people in the video. The camera then detects that skin tone in the images and modifies the images to obscure the facial information. Fidaleo et al. [FNT04] propose to use a filter for detecting and removing the facial information before saving it onto the server until the person's behavior is considered suspect. Kitahara et al. [KKH04] detect and pixelize the facial region to protect privacy in video recorded by mobile cameras. Similarly, Newton et al. [NSM05] replace the faces in the images with eigen-faces so that the face recognition software fails. The work mainly deals with the images in which a high resolution frontal face is present, which is generally not true for surveillance data. Boult [Bou05] uses encryption to obscure privacy regions which can be inverted later by authorized person using decryption key. The privacy regions are determined using face detection. Chattopadhyay and Boult [CB07] extend this work to implement the proposed privacy framework on a Blackfin DSP architecture (PrivacyCam). Martinez-ponte et al. [MPDMD05] use face detection and tracking to detect the privacy regions and move them to the lowest quality layer of JPEG 2000. Brassil [Bra05, Bra09] proposes to use location sensitive mobile devices (e.g. GPS receiver) to protect the individual's privacy.

Using mobile device, people can express their preference for privacy preservation via wireless communication such as GPRS. The segment containing particular person is processed to remove privacy information by detecting and obscuring the faces. Chen et al. [CCYY07] detect faces from the medical images and ask users to label humans. The images are further obscured to hide human appearance. Chaudhari et al. [CCV07] detect and block (replace by a colored box) faces to protect privacy. To reduce privacy loss through audio, the authors propose a pitch shifting algorithm. Carrillo et al. [CKM08] detect and encrypt the faces to protect privacy. Dufaux et al. [Duf11] scramble facial region to protect privacy. Different scrambling methods are compared based on the rate distortion (PSNR) with and without scrambling. Schiff et al. [SMM+09] use visual markers (colored hats) to localize faces and obscure them to protect anonymity of the individuals. All of the above described works assume that if the face is obfuscated in the video, it is zero privacy loss, if the face is visible, it is full privacy loss:

$$\Gamma = \begin{cases} 1 & \text{if the face is recognizable;} \\ 0 & \text{otherwise.} \end{cases} \tag{2.3}$$

Most of these works have put the main emphasis on the robustness of the privacy protection, the utility of the data is generally ignored. They do not provide any computational models to measure the utility and the privacy of the video data. The works are limited to a single camera video and they do not analyze the impact of multiple events on the identity leakage. Later we will see that the privacy models described in Equation 2.2 and 2.3 are not robust enough. The identity loss can still occur even when these equations predict zero privacy loss. In 2009, Babaguchi et al. [BKUT09] conducted a user study about people's sense of privacy. In the study they found that users felt least privacy violation when their body was replaced by a text annotation in the images compared to other images which showed original image, face removed image, silhouette transformed, an image with human body replaced with age and gender as text annotation. It further shows that it is not only the identity, but the association of the identity with the sensitive information that causes privacy loss.

2.1.3 Summary

In the past works, the privacy loss is often viewed as a set of predefined discrete values. This set could be of size two (privacy is preserved or lost) [TLB+06, KTB06, SPH+05, PCH09, Qur09] or a fixed number [MVW08]. Further, it has been observed that researchers have focused only on the *who* aspect (face information) and they have overlooked other implicit inference channels associated with *what* (activity), *where* (location where video is recorded) and *when* (time when video is recorded). An adversary can observe the behavior, look at the places visited and use prior knowledge to infer the identity information. To the best of our knowledge, we are the first to model the privacy loss as a continuous variable in the range, considering both explicit and implicit channels.

Table 2.1 presents a summary of existing works and shows how the proposed work is novel compared to them. This summary has been provided from the following different aspects: whether implicit inference channels are considered; whether privacy loss is modeled for surveillance video from multiple cameras; whether the notion of sensitive information has been used in privacy loss computation; whether privacy loss is determined as a binary or continuous value; whether privacy loss is determined based on single or multiple events in the video? It is clearly shown in the table that the proposed work is novel in many aspects.

2.2 Data Transformation

Once the characteristics of video are identified that cause privacy loss such as image regions or event sequences, the next step is to transform the video data such that the privacy can be protected. One trivial solution to this problem is to remove everything from the images, but such video has no utility. For example, a video captured for activity monitoring should serve the utility of activity detection while preserving the privacy. In order to study this tradeoff between the privacy and the utility, we need to have computational models of both. In Table 2.2, we present a comparison of our proposed work with other works with respect to the following points: whether implicit identity leakage channels (e.g. location, time and activity information) have been used for assessing the privacy loss; whether privacy loss is modeled as a binary value (1 and 0), a set of fixed values, or a continuous function; whether a tradeoff between the

Table 2.1: A Summary of Related Work for Privacy Modeling.

Work	Implicit Channels	Multi-Camera	Sensitive Info (SI)/ Identity (I)	Modeling Binary/ Continuous	Consider-ation of Events
Ackerman et al. [AS95]	No	No	SI	Binary	No
Hudson and Smith [HS96]	No	No	I	Binary	No
NYNEX [LGS97]	No	No	SI	Binary	No
Lee et al. [LSG97]	No	No	SI	Binary	No
Zhao and Stasko [ZS98]	No	No	SI	Binary	No
Berger [Ber00]	No	No	I	Binary	No
Tansuriyavong and Hanaki [TH01]	No	No	SI	Binary	No
Al-Muhtadi et al. [AMCK+02]	No	No	SI	Binary	No
Campbell et al. [CAMN+03]	No	No	SI	Binary	No
Beresford and Stajano [BS03]	No	No	SI	Binary	No
Kitahara et al. [KKH04]	No	No	I	Binary	No
Fidaleo et al. [FNT04]	No	No	I	Binary	No
Wickramasuriya et al. [WDMV04]	No	No	I	Fixed levels	No
Senior et al. [SPH+05]	No	No	I	Binary	No
Newton et al. [NSM05]	No	No	I	Binary	No
Boult [Bou05]	No	No	I	Binary	No
Martinez-ponte et al. [MPDMD05]	No	No	I	Binary	No
Zhang et al. [ZCC05]	No	No	I	Binary	No
Brassil et al. [Bra09]	No	No	I	Binary	No
Koshimizu et al. [KTB06]	No	No	I	Binary	No
Spindler et al. [SWH+06]	No	No	I	Fixed levels	No
Thuraisingham et al. [TLB+06]	No	No	I	Binary	No
Bhaskar and Ahamed [BA07]	No	No	SI	Binary	No
Chen et al. [CCYY07]	No	No	I	Binary	No
Chaudhari et al. [CCV07]	No	No	I	Binary	No
Carrillo et al. [CKM08]	No	No	I	Binary	No
Moncrieff et al. [MVW08]	No	No	I	Fixed levels	No
Paruchuri et al. [PCH09]	No	No	I	Binary	No
Qureshi et al. [Qur09]	No	No	I	Binary	No
Cheung et al. [CVP+09]	No	No	I	Binary	No
Schiff et al. [SMM+09]	No	No	I	Binary	No
Bagues et al. [BZK+10]	No	No	SI	Binary	No
Dehghantanha et al. [DUM10]	No	No	SI	Binary	No
Zhu et al. [ZCZM10]	No	No	SI	Binary	No
Dufaux et al. [Duf11]	No	No	I	Binary	No
Saini et al. [SAM+10]	Yes	No	I & SI	Continuous	Single
Proposed Model	**Yes**	**Yes**	**I & SI**	**Continuous**	**Multiple**

privacy loss and the visual distortion of the whole frame due to data transformation (we call it *utility loss*)has been examined; and which of the approaches (selective obfuscation or global operations) has been adopted.

Most researchers [BEG00],[SPH+05], [FNT04], [WDMV04], [KTB06], [TLB+06], [CKM08], [PCH09], [Qur09] have used selective obfuscation to preserve the privacy in the surveillance videos. They have adopted the traditional approach, which is to detect the region of interest (e.g. face or blob) and hide it. Since this approach is limited by the accuracy of the detectors, privacy cannot be guaranteed. The other set of works do not rely on the detectors and go for global transformation of the whole image [AS95, LGS97, LSG97, BEG00]. In these works, the obfuscation function (bluring, quantization, pixelization etc.) is applied on the whole image to hide the privacy information. This approach is too pessimistic and affects the utility of the data adversely.

In a recent survey, Chinomi et al. [CNIB08] compare different methods for obscuring people in the video data. In PriSurv [CNIB08], the appearance of a person is manipulated depending on the viewer. The transformations are shown in Figure 2.5. The images are arranged in decreasing order of the privacy loss. Following transformations are explored:

- As-Is (Figure 2.5.a)

  In this transformation the video is kept in its original form. Effectively, the video contains all the visual information that can help the adversary to learn the identity of the individuals as well as the sensitive information about them. Therefore, this forms the lowest level of privacy protection. Figure 2.5.a shows the original image as a result of this transformation.

- See-through (Figure 2.5.b)

  In see-through transformation the pixel values of the foreground and the background are blended such that the background is visible through the object. The viewer cannot obtain information from the video as precisely as from the original. Hence, it provides better privacy than showing the original image, but this improvement is not significant as most of the visual information is still accessible and adversary can easily obtain the identity information.

- Monotone (Figure 2.5.c)

  Some people might be sensitive to the color related information e.g. color of clothes, skin tone etc. To preserve privacy in these scenarios, the color information of the image is removed in 'monotone' transformation. This transformation is reliable as it does not involve any object detection and minimally affects the task of video surveillance. However, the transformation is effective only if the individuals in the video are only sensitive to color, because other information is still available in the video. This transformation is also not effective for hiding the identity.

- Blur (Figure 2.5.d)

  In this transformation the object is blurred so that the sharp details of the object are hidden. Enough amount of blurring can hide the identity of the individuals, but it will also affect the quality of surveillance if the motion information is not preserved. This method depends on the accuracy of the object detector, and it will fails when the detector fails. However, in worst cases we can blur the whole image globally.

- Pixelization (Figure 2.5.e)

  Pixelization is very effective technique in hiding the appearance of humans in the video if the object can be detected accurately. As can be seen in the Figure 2.5.e, it is very difficult to identify the individuals in a pixelized image. However, it distorts the object boundaries more severely than blurring. If the boundary information is important for the intended surveillance task, this transformation is not good.

- Edge (Figure 2.5.f)

  In this transformation, edge detection is performed on the object and the object is replaced by extracted edges in one color. In the Figure 2.5.f we can see that this method is effective in hiding the identity information and appearance details from the image. The challenge in this method is accurate detection of the object boundary and nature of background. If the background has high frequency information, inaccurate object detection will distort the shape of object making surveillance very difficult.

- Border (Figure 2.5.g)

  If the object can be detected accurately, it can be replaced by object boundary which is

sufficient for determining the activity information. However, there are mainly two problems with this approach: (1) it is very hard to accurately determine the object boundary (2) if the individual is carrying an auxiliary object that is security threat, this transformation completely eliminates this information. Further, it is not possible to globalize this transformation.

- Silhouette (Figure 2.5.h)

  Replacing an object by its silhouette is similar to replacing by the border. The identity of the individual in the video is preserved effectively but the method depends on the accuracy of silhouette detection method. In this case also we do not have the option of global transformation as that will eliminate everything from the image.

- Box (Figure 2.5.i)

  Replacement of the object by a painted box hides all details of the object except the height and width. While this method is also limited by the accuracy of the object detectors, it can affect the surveillance more adversely. The security threats due to image regions in the box cannot be detected.

- Bar (Figure 2.5.j)

  This transformation is one step further of the box transformation. In this transformation, the width information is also removed by replacing the object by a single line. With this transformation, only few surveillance tasks like people counting, crowd flow, intrusion etc. can be performed. Given the inaccuracies of the object detectors, this method is highly unreliable from privacy perspective.

- Dot (Figure 2.5.k)

  In this transformation, the object is replaced by a single dot hiding all details of the object but the location. No activity detection can be performed on the video which is transformed with this method. This transformation can be used in scenarios where people counting is the only surveillance task. Object detection also limits applicability of this method in real systems.

- Transparency (Figure 2.5.l)

  This transformation is the other extreme of the 'As-Is' transformation. In this transfor-

mation, the object is completely removed as if the object was not there. This is highest level of the privacy that can be provided. This method does not require accurate object detectors as all the frames can be simply replaced by the static background frame. No surveillance can be performed after this transformation.

As shown in Table 2.2, our work is different from the works of other researchers in many aspects. First, we examine the implicit identity leakage channels which have been ignored in the past. Second, in our work the privacy loss is modeled as a continuous variable compared to binary or a predefined set of fixed values. Third, the proposed privacy preserving method presents a tradeoff between the utility and the privacy for the data publication scenario. Finally, the proposed method examines a hybrid approach for data transformation. In our approach, we propose to use combination of quantization and blurring that achieves improved tradeoff between the privacy and the utility.

2.3 Data Publication

Publication of video data is very useful for many user communities. Many applications related to ethnography, psychology and policy making can benefit considerably from analysis of this data. For example, researchers working in the field of automated video surveillance can test their algorithms on the published video. There are few video datasets available for public download and testing. The Honda/UCSD Video dataset [LHYK03, LHYK05] contains video sequences for evaluating face tracking/recognition algorithms. The dataset contains videos clips contributed by volunteers and Youtube videos. The videos contained many normal life scenarios (concepts) like birthday parties and weddings. There are few datasets for testing human action recognition algorithms which consist of video clips from Hollywood movies [MLS09, LMSR08]. Kodak's consumer video dataset can be used for semantic indexing of the videos. NIST [NIS10] has provided broadcast news video dataset for researchers in information retrieval field. These videos are artificially generated except when the users voluntarily donate the videos. They also include dataset for surveillance event detection. PETS [PET11] has provided datasets for the evaluation of vision algorithms deployed in video surveillance systems such as human tracking, crowd analysis, left baggage detection, vehicle tracking etc.

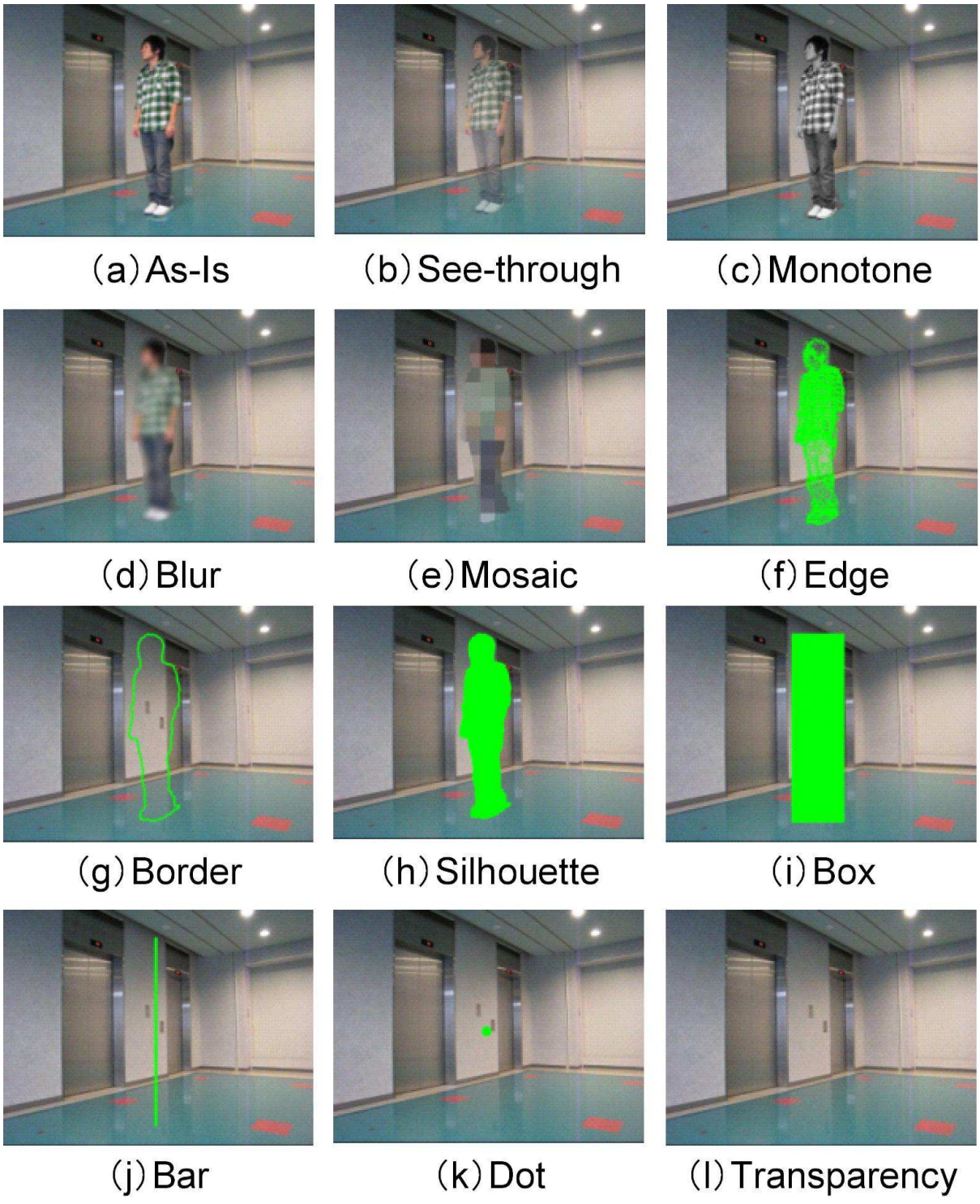


Figure 2.5: Different data transformations explored for privacy protection by Chinomi et al. [CNIB08]. Image taken from [CNIB08]

Table 2.2: A comparison of the proposed work with the existing works on privacy-aware surveillance

The work	Identity leakage channels used	Utility quantified?	Approach adopted	Global Obfuscation (GO)/Selective obfuscation(SO)
Ackerman et al. [AS95]	No	No	Iconic representation	GO
Hudson and Smith [HS96]	No	No	Iconic representation	SO
NYNEX [LGS97]	No	No	Image Transformation	GO
Lee et al. [LSG97]	No	No	Iconic representation	SO
Zhao and Stasko [ZS98]	No	No	Image transformation	GO
Boyle et al. [BEG00]	Explicit	No	Image transformation	GO
Berger [Ber00]	No	No	Face obfuscation	SO
Tansuriyavong and Hanaki [TH01]	No	No	Silhouette obfuscation	SO
Kitahara et al. [KKH04]	No	No	Face obfuscation	SO
Fidaleo et al. [FNT04]	Explicit	No	Face obfuscation	SO
Wickramasuriya et al. [WDMV04]	Explicit	No	Blob obfuscation	SO
Senior et al. [SPH+05]	Explicit	No	Face and blob obfuscation	SO
Newton et al. [NSM05]	No	No	Face obfuscation	SO
Boult [Bou05]	No	No	Face encryption	SO
Martinez-ponte et al. [MPDMD05]	No	No	Face compression	SO
Zhang et al. [ZCC05]	No	No	Blob obfuscation	SO
Brassil et al. [Bra09]	No	No	Face obfuscation	SO
Koshimizu et al. [KTB06]	Explicit	No	Silhouette obfuscation	SO
Spindler et al. [SWH+06]	Explicit	No	Blob obfuscation	SO
Thuraisingham et al. [TLB+06]	Explicit	No	Face obfuscation	SO
Chen et al. [CCYY07]	No	No	Face obfuscation	SO
Chaudhari et al. [CCV07]	No	No	Face obfuscation	SO
Carrillo et al. [CKM08]	Explicit	No	Face obfuscation	SO
Moncrieff et al. [MVW08]	Explicit	No	Face obfuscation	SO
Dufaux et al. [Duf11]	No	Yes	Blob obfuscation	SO
Paruchuri et al. [PCH09]	Explicit	No	Blob obfuscation	SO
Qureshi et al. [Qur09]	Explicit	No	Blob obfuscation	SO
Cheung et al. [CVP+09]	No	No	Blob obfuscation	SO
Schiff et al. [SMM+09]	No	No	Face obfuscation	SO
Proposed work	**Explicit and implicit**	**Yes**	**Data transformation**	**GO**

The above mentioned datasets are either staged or taken from movies. The videos generally contain scenes created by the actors in constrained scenarios. Due to limited size, they generally do not capture the scenarios in entirety. In contrast, surveillance videos are virtually infinite in size and represent the real scenarios an algorithm may encounter when deployed in systems. Also, the surveillance videos can be used for various social studies involving human behavior such as video ethnography, which is not possible with the available video datasets as the events and activities are generally artificially generated.

Activity analysis is the basic building block of many video analysis applications; therefore, our goal is to publish video data in such a way that not only the utility of the published data is maintained, but the privacy loss is also minimized. To achieve this goal, contrary to the past works in which the approach has been to detect and obfuscate the facial region in the images, we propose to transform the video data. The data transformation function is chosen such that face, location, and time cannot be detected; however, the detectors that contribute to the utility would work.

2.4 Privacy in Statistical Data Publication

In the statistical data publication, the concept of identity preservation is well studied. The statistical data is published as tuples which consist of the following fields:

- Direct identifiers: These attributes almost uniquely identify a person. Examples include name, ssn, passport number, email id, address, and registration number. The direct identifiers are removed from the tuples before publishing.

- Quasi identifiers: These identifiers relate the data to a group of people. Examples are age, school name, postal code, country, height, weight, color, profession, etc. The application of published data determines which quasi-identifiers to publish.

- Sensitive attributes: These are mainly the attributes people do not want others to know. For example, the hospital records have disease description which people generally do not want to disclose to others. The mapping from identities to these sensitive attributes results in privacy violation. Privacy protection is achieved by releasing the data in such a form that this mapping cannot be established by the adversaries.

The statistical data is *anonymized* before publishing in order to ensure privacy [FWCY10]. The traditional methods of anonymizing data are removal of direct identifiers; and perturbation and sub-sampling of quasi identifiers. In all these cases, there is one entity which is considered 'trusted' and data sanitization is performed by that entity. Later it is found that even after primitive sanitization, the privacy preservation is not guaranteed [Swe02]. This lead researchers to explore other techniques to generalize the data to sustain the privacy attacks by an adversary with auxiliary knowledge. These include k-anonymity [Swe02], l-diversity [MKGV07], t-closeness [LL07], and δ- presence [NAC07]. A recent work [Dwo06] argues that it is impossible to achieve the privacy in absolute sense and introduced differential privacy. A detailed summary of these techniques can be found in [FWCY10].

2.5 Summary

In this chapter we have reviewed previous works on privacy modeling, data transformation, and video data publication. We have found that the current models ignore the identity leakage from implicit channels. Therefore, the current models of privacy are not robust and only give false sense of privacy protection. There is a need to extend the privacy models to consolidate the identity leakage from both implicit and explicit channels. The data transformation functions proposed in the literature are limited by the accuracy of the vision detectors. Therefore, they provide unreliable privacy protection. Further, we need continuous models of the privacy and the utility in order to study tradeoff between the two, which are not available in the literature. Publicly available datasets are very useful for many research communities. The current datasets are either staged or movie clips and lack realness. The enormous amount of surveillance footage captures real events, but it cannot be published due to privacy concerns. Therefore, there is a need to explore privacy preserving publication of surveillance video data.

Chapter 3

Privacy Model for Single Camera Video

Video is a rich source of information. The first step towards privacy preserving use of video data is privacy modeling. Surveillance systems record video using multiple cameras and extract the embedded information for the assessment of the situation [DV08]. Similarly, other scientific and social studies can benefit from the availability of video data. The biggest problem with the use of video data is a privacy violation [KTB06]. People do not like their activities being recorded and watched by others. The challenge here is to minimize the privacy violation which might occur due to public access of the video data while still preserving its usability. A detailed modeling of the privacy loss may expose various leakage channels through which the privacy violation occurs, which can be blocked to reduce the same.

Traditionally, it has been assumed that presence or absence of the facial information determines the privacy loss. Although this method of assessing privacy counters the *explicit* inference channel of identifying people by their facial information, it often leaves various *implicit* inference channels that can be used by adversaries in order to infer the identities of individuals in the video. In the proposed model, the privacy loss is determined based on the *explicit* as well as *implicit* identity leakage channels. The following are examples of implicit inference channels that can cause identity leakage of the individuals in the video:

- *What* are the activities recorded in the video? This can be learned and associated with a person or group of people. For example the way people greet others can be specific for

individuals.

- *Where* is the video recorded? Spatial clues can help in determining a person's identity. For example landmarks and text legends can reveal the location and limit anonymity to a small number of people.

- *When* is the video recorded? The time information can further reduce the ambiguity in the identity. For example, a person entering a shop in the early morning is most likely to be one of the employees.

We applied the privacy model to transform video in data publication scenario that simulates the worst case situation of the privacy loss in video surveillance systems as the data is accessible to everyone. For preserving an individual's privacy, we need to transform data into such a form that no identity information can be inferred from the transformed data. However, the process of data transformation may affect the utility of the data. The core idea of our approach is to transform the video data in such a manner that minimizes the utility loss and the privacy loss simultaneously. To measure the utility loss, we determine the important tasks required for the given application and propose a task based model to quantify the utility loss between zero and one.

We summarize the main contributions of this chapter as follows:

- We first measure the identity leakage through implicit and explicit inference channels as degree of anonymity; and then compute the privacy loss. To the best of our knowledge, this is the first work that models the privacy loss as a continuous variable for video data publication scenario.

- We propose a task based utility model to study the tradeoff between the utility and the privacy. To the best of our knowledge, this is the first attempt to study the tradeoff between the privacy and the utility over a continuous scale.

- We investigate a suitable transformation function that minimizes the utility loss as well as the privacy loss of the published data.

3.1 Chapter Organization

We start the chapter with important terms and definition (Section 3.2). The privacy is modeled (Section 3.3) as a product of the identity leakage and the sensitivity index. While the identity leakage is determined based on both implicit and explicit channels (Section 3.3.1), the sensitivity index captures the possibility of the video containing the sensitive information (Section 3.3.2). For data publication (Section 3.4), a problem is formulated (Section 3.4.1) to study tradeoff between the utility loss (Section 3.4.2) and the privacy loss for different data transformation methods (Section 3.4.3); followed by experimental results (Section 3.4.4). The research findings of the work presented in the chapter are summarized in Section 3.5.

3.2 Definitions

Following are some of definitions and the key terms used throughout the chapter.

Definition 7 ***Event****: The event definition has been adopted from [AKJ06]: 'Event is a physical reality that consists of one or more living or non-living real world objects (who) having one or more attributes (of type) being involved in one or more activities (what) at a location (where) over a period of time (when)'.*

Definition 8 ***Evidence****: This is the information which is extracted from the video data. The facial information is termed as "who" evidence, any activity related information is termed "what" evidence, temporal information is said to be "when" evidence, and spatial information as "where" evidence. These evidence types have been identified as main aspects of a generic event model [WJ07].*

Definition 9 Privacy Loss*: This is the certainty with which an adversary is able to gain the sensitive information about an individual in the video.*

Definition 10 Utility Loss*: Utility loss of the published video data refers to the decrease in the degree of accuracy by which analysis tasks can be accomplished with respect to the original data.*

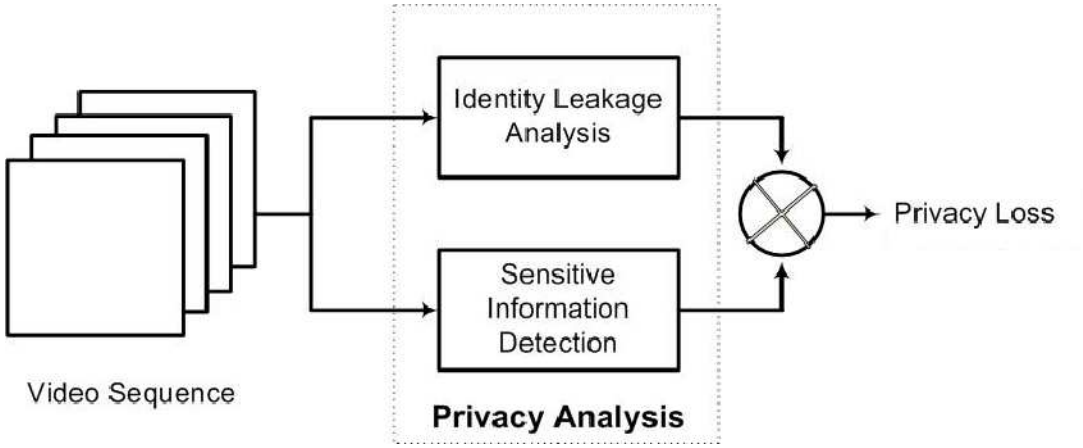


Figure 3.1: Assessment of the privacy loss of the individuals in the video. The privacy loss is determined based on the identity leakage and associating the identity with the sensitive information present in the video.

3.3 Proposed Privacy Model

In the past works, it has been assumed that the privacy loss is equivalent to either the identity leakage [WDMV04], [CKM08] or the sensitive information present in the video [AS95], [LGS97], [ZCZM10]. We recognize that the privacy loss occurs when the adversary is able to map an identity to the sensitive information in the video, for example their habits, physique, companions etc. With this observation the privacy loss is modeled as a combined function of identity leakage and sensitivity index (a measure of the sensitive information in the video). The block diagram of the proposed privacy analysis framework is shown in Figure 3.1. We first calculate the identity leakage and the sensitivity index separately and then combine them to calculate overall privacy loss. The adversary can either be a human being with prior knowledge or an automated system with pattern information obtained through data mining and similar learning techniques.

3.3.1 Identity Leakage

Different forms of the identity leakage of the people in the video are based on four types of evidences: *who* a person is, *what* activity the person is performing, *when* the person is doing that activity, and *where* is the person doing it. Except *what* evidence, all other evidences could be detected in a single frame of the video. The identity leakage from individual evidences or group of evidences are modeled as a measure of anonymity, which is popularly used in statistical data publication to measure the privacy loss [Swe02], [MKGV07], [LL07] and refers to the state

of being anonymous. For evidences *what, when,* and *where*, the anonymity is calculated directly, whereas for the *who* evidence, it is estimated to be close to unity.

The proposed model consolidates the identity leakage from both implicit and explicit channels to provide more robust measures of privacy loss. We will use notations wo, wt, wn, and wr to refer to the detection of *who, what, when*, and *where* evidences. The explicit identity leakage I_{ex} occurs due to the presence of *who* evidence and it is modeled as I_{wo} i.e. $I_{ex} = I_{wo}$, where I_{wo} is the identity leakage due to *who* evidence.

The identity leakage due to implicit channels depends on the number of evidences detected. When there are no people in the frame, the identity leakage is zero. The identity leakage only occurs when one or more people are present in the video. Since people are always involved in some activities, *what* evidence is always detected whenever there are people in the video. With *what* evidence, we can also have *when* and *where* evidences detected. Therefore, the implicit identity leakage I_{im} is modeled as follows:

$$I_{im} = \begin{cases} 0 & \text{if no people are present;} \\ I_{wt} & \text{if only } \textit{what} \text{ detected;} \\ I_{wt,wr} & \text{if } \textit{what} \text{ and } \textit{where} \text{ detected;} \\ I_{wt,wn} & \text{if } \textit{what} \text{ and } \textit{when} \text{ detected;} \\ I_{wt,wn,wr} & \text{if } \textit{what, when} \text{ and } \textit{where} \\ & \text{detected.} \end{cases} \tag{3.1}$$

The detailed modeling of I_{wo}, I_{wt}, $I_{wt,wn}$, $I_{wt,wr}$ and $I_{wt,wn,wr}$ is described below.

- I_{wo}: The evidence *who* almost uniquely identifies the person. It is generally assumed that everyone has unique facial features that are used to distinguish that person from others; therefore, if the face is detected in a video, providing *who* evidence, the identity leakage is unity as shown in Equation 3.2.

$$I_{wo} = \begin{cases} 1 & \text{if the face is recognizable;} \\ 0 & \text{otherwise.} \end{cases} \tag{3.2}$$

- I_{wt}: Individually, this evidence contributes negligibly to the identity leakage. Because lots of people do similar activities in the world, even if we recognize the activity correctly, the

person will be identified as one from a large group of people. For example, we recognize that a person in the video leans while shaking hands. If *when* and *where* evidences are not present, there are thousands of people in the world who lean while shaking hands; hence the identity leakage is very small. Since the identity leakage is inverse of the associated group size, it is a very small value that can be ignored in most cases. The value of I_{wt} is calculated as follows:

$$I_{wt} = \frac{\rho}{G_{wt}} \tag{3.3}$$

where ρ is the number of people simultaneously present in a frame, G_{wt} is the average number of people per activity detected in the video, and $\rho \leq G_{wt}$. To calculate this value, we define a polymorphic function $\mathcal{H}$ that returns the number of people satisfying the conditions provided as arguments. Now, G_{wt} is calculated as:

$$G_{wt} = \frac{1}{n_1} \sum_{i_1=1}^{n_1} \mathcal{H}(\xi_{i_1}) \tag{3.4}$$

where n_1 is the number of activities, ξ_{i_1} is i_1^{th} activity.

- $I_{wt,wn}$: It is the identity leakage when the people are present in the video, and the time information is also available. The identity leakage due to these two evidences is calculated by determining the anonymity. The total time (24 hours in our case) is divided into different slots and the number of people performing particular activity in each time slot Δt is counted. The average number of people in a time slot is treated as the group size $G_{wt,wn}$, which provides the measure of anonymity:

$$G_{wt,wn} = \frac{1}{n_1 * n_2} \sum_{i_1=1}^{n_1} \sum_{i_2=1}^{n_2} \mathcal{H}(\xi_{i_1}, \Delta t_{i_2}) \tag{3.5}$$

where n_2 is the number of total time slots. The identity leakage $I_{wt,wn}$ is calculated as:

$$I_{wt,wn} = \frac{\rho}{G_{wt,wn}} \tag{3.6}$$

where, ρ is the number of people detected simultaneously in the frame.

- $I_{wt,wr}$: The availability of *what* and *where* evidence results in $I_{wt,wr}$ identity leakage. The *where* evidence is generally made available by detecting landmarks or text legends in the video. The presence of this evidence associates a person to a group of people who are usually found at the place of recording. Similar to $I_{wt,wr}$, firstly the group size for each popular text legend or landmark is calculated; then the average group size $G_{wt,wr}$ is calculated as a measure of anonymity:

$$G_{wt,wr} = \frac{1}{n_1 * n_3} \sum_{i_1=1}^{n_1} \sum_{i_3=1}^{n_3} \mathcal{H}(\xi_{i_1}, \lambda_{i_3}) \tag{3.7}$$

  where n_3 is the number location identifiers and λ_i is i^{th} location identifier. Once we determine the association group size $G_{wt,wr}$, the identity leakage can be calculated by dividing the number of people detected ρ by the group size:

$$I_{wt,wr} = \frac{\rho}{G_{wt,wr}} \tag{3.8}$$

- $I_{wt,wn,wr}$: The presence of all three type of evidences results in significant increase in the identity leakage, and the individuals in the video can be associated to a small group of people. For example, lots of people snack; lots of people do lots of things at 4 pm; but John and Mike are amongst the only persons at Great Residences who snack at 4 pm. So three pieces of evidence, snacking (*what*), Great Residences (*where*), and 4 pm (*when*) reduce the group size to 2 leading to the identity leakage of 0.5, assuming there is only one person in the current frame. Hence, with all three type of evidences present in the video, the identity leakage is calculated as:

$$I_{wt,wn,wr} = \frac{\rho}{G_{wt,wn,wr}} \tag{3.9}$$

  where $I_{wt,wn,wr}$ is the identity leakage due to simultaneous present of *what, when*, and *where* and $G_{wt,wn,wr}$ is the associated average group size which is calculated as follows:

$$G_{wt,wn,wr} = \frac{1}{n_1 * n_2 * n_3} \sum_{i_1=1}^{n_1} \sum_{i_2=1}^{n_2} \sum_{i_3=1}^{n_3} \mathcal{H}(\xi_{i_1}, \Delta t_{i_2}, \lambda_{i_3}) \tag{3.10}$$

While it is generally difficult to accurately measure group sizes (G), they can be estimated with the help of domain knowledge. For example, the group size at a personal office can be in a single digit. On the other hand, the group size for a subway station can be in thousands.

Overall Identity Leakage

When multiple evidences are detected simultaneously, the overall identity leakage from each frame is calculated as the maximum of the implicit and explicit identity leakages i.e.

$$I_f = MAX\{ \underbrace{I_{ex}}_{explicit} , \underbrace{I_{im}}_{implicit} \} \tag{3.11}$$

where I_f is the identity leakage from a frame. While the identity leakage is calculated for each frame of the video, the frame which causes the worst case identity leakage is considered the representative of the whole video for both explicit (I_{ex}) and implicit (I_{im}) channels. Let I be the identity leakage for the video clip under consideration, it can be determined as follows:

$$I = MAX\{I_f \mid \forall f \in V\} \tag{3.12}$$

In Equation 3.11 and 3.12 we consider the worst case scenarios to emphasize that the privacy preserving techniques need to be robust enough to block these channels in every frame of the video. If the person in the video is recognized even in one frame, this identity information can be propagated to other frames and cause the privacy loss.

3.3.2 Sensitivity Index

The privacy loss and the identity leakage are two separate phenomenon. Mere identity leakage does not generally cause privacy loss. For example, a video which only shows full frontal face reveals the identity of the person quite accurately. However, if no other information can be learned from video (activity, place, time, etc.), people generally do not feel it to be a privacy loss. It is worth mentioning that a person's face is usually public and can be easily accessed in many other ways from the web [R.08]. On the other hand, if the video also shows which place the person is visiting or whom the person is meeting, it might be a privacy loss for some individuals. Similar situation can be found in the statistical data publication where well structured data

Table 3.1: Commonly found sensitive information.

Sensitive Attribute	Example
Activity	Showing middle finger when alone.
Spatial Information	Generally we do not want strangers to know which places we visit.
Time	Some people mind when others associate their activities with timing patterns.
Gesture	People make strange gestures while they are alone and do not want others to watch that.
Clothes	Many teens wear clothes which they do not want their parents to know.
Physique	People with atypical physique may be sensitive to that e.g. height.
Habits	Most people have some personal idiosyncratic sensitive habits like twiddling fingers under stress.
Companion Information	Some people do not want everyone to know whom they associate with.
Associated Objects	What we carry with us.

records of individuals are published after removal of the direct identifiers [FWCY10]. There, the privacy loss occurs when an adversary is able to map the identity to the sensitive information, stored in the sensitive information fields of the published data records. For example, medical data records might contain disease names as sensitive information.

Privacy loss occurs when the adversary is able to associate the identity with the *sensitive* information. Video generally contains enormous amount of information which might qualify as *sensitive* information. Which information is sensitive and which is normal depends on the individuals and it may vary from person to person [Lan01]. Yet, Table 3.1 enumerates commonly found video attributes which are considered sensitive. Fortunately, most of these attributes are well captured by the evidence types *what, when*, and *where*, and the identity leakage through these implicit channels can provide a basis to determine the privacy loss.

To calculate the privacy loss, we need to obtain not only the identity leakage but also the amount of the sensitive information in the video. We assume that the information in the video consists of a set of attributes and some of these represent the sensitive information. Let $A = \{a_1, a_2, ...a_l\}$ be the set of attributes that can potentially be sensitive information. Let W be priority vector defined as:

$$\mathcal{W} = \{w_k \mid k \in [1, l], w_k \geq 0, w_1 + w_2, ... + w_l = 1\} \tag{3.13}$$

The elements of the vector are weights (w_k) which are set by the individuals seen in the video and

they reflect their priority of the corresponding attribute as sensitive information. By analyzing the video it can be determined what sensitive attributes are detected in the video:

$$S = \{s_k \mid k \in [1, l]\} \tag{3.14}$$

The elements of the vector are determined as follows:

$$s_k = \begin{cases} 1 & \text{if } k^{th} \text{ attribute is detected;} \\ 0 & \text{otherwise.} \end{cases}$$

The sensitivity index is calculated as a dot product of the priority vector and the sensitivity vector:

$$\Psi = \mathcal{W}.S \tag{3.15}$$

The equation reflects that any information in the video only adds to the privacy loss if it is also sensitive to the individuals in the video.

3.3.3 Privacy Loss

If V is the video data under consideration, the privacy loss can be defined as follows:

Definition 11 *The Privacy loss due to published data V is represented by $0 \leq \Gamma(V) \leq 1$. $\Gamma = 0$ implies no privacy loss and $\Gamma = 1$ represents the worst case where the individual's identity, along with other information such as activity, time and place, can be determined exactly.*

If we remove the identity information completely, the video cannot cause privacy loss to any individual; no matter how much sensitive information the video contains. This is because the sensitive information cannot be associated to anyone. Similarly, if there is nothing sensitive in the video, it generally does not cause privacy loss. In both the cases, the resulting privacy loss is zero. Hence, the privacy loss can be calculated as the product of the identity leakage and the sensitivity index.

$$\Gamma = I \times \Psi \tag{3.16}$$

The implication of the equation is that both the sensitive information and the identity leakage can be manipulated to control the privacy loss.

3.3.4 Absence Privacy

Other than the privacy loss of the people who are present in the video, the background information also causes privacy loss to the people who are absent in the video. Absence information has been recently recognized as a sensitive information in many scenarios [FRVM+10]. Research in privacy-aware use of video data has been focused on the privacy loss that occurs when there are people present in the video; however, there are no attempts in literature to measure the privacy loss that occurs due to the absence of people in a video. Even a blank video with only background information enables the viewers to gain additional information about the absence of individuals at the location where the camera is placed. This additional information was not available to the viewer before watching the video; therefore it may lead to privacy loss [Dal77]. The following examples demonstrate the impact of such a privacy loss:

- Consider a company campus where there are two clinics: general clinic and X disease clinic. There are no surveillance cameras around X disease clinic to protect the privacy, but the area around general clinic is monitored. John tells his friends that he is going to hospital but no one is seen in cameras around general clinic on that particular day. Traditional privacy loss models would suggest zero privacy loss in this case; however, it compromises the privacy of John who does not want others to know his/her disease.

- In social networks such as Facebook, let us consider that someone uploads a video clip of a friend's Birthday party. If people having access to this video clip find that John should have been there but is not present in the video, it can provide some evidence that John might not have joined this party. This can lead to privacy loss for John.

Intuitively, the identity leakage gives us the probability of someone being present in the video. Therefore, the absence of a person can be modeled as a compliment of the identity leakage and the absence privacy loss (Γ') could be calculated as follows:

$$\Lambda = (1 - I) \times \Psi \tag{3.17}$$

As discussed earlier, the privacy loss occurs when an adversary is able to learn the sensitive information about an individual from the exposed video that could not be learned without it [Dwo06]. The privacy loss in a video can be of two types: 1) *privacy loss due to the presence of people*, e.g. an adversary is able to determine that the person was present at a particular place and gain additional sensitive information about that person, such as place, time, activity, accompanying people etc. (Equation 3.16); and 2) *privacy loss due to absence of people*, e.g. adversary could know that the person was not present at a particular place ((Equation 3.17)). In the second case, if the person was expected to be there, but s/he was not there, it might cause privacy loss. Which of these two privacy losses is more important, depends on the scenario. For example, if it is a hospital or police station, people do not want to be observed there since their presence can leak some sensitive information about them. On the other hand, if it is an office scenario, people worry if others know that they did not reach the office on time, which can also cause privacy loss.

3.4 Privacy-Aware publishing of Surveillance Video

Publication of video data is very useful for many user communities. Many applications related to ethnography, psychology and policy making can benefit considerably from availability of natural video data. Further, the researchers working in the field of automated video surveillance can test their algorithms on the published video. There are few video datasets available for public download and testing [SOK06], [LHYK03]; however, these videos are generally staged or movie clips with only limited scenarios. The huge amounts of video recorded by CCTV surveillance systems capture real scenarios in their entirety. This data is generally discarded after some time due to obvious privacy concerns [SWH$^+$06], [Lev06]. In this section, we extensively investigate the privacy issues associated with surveillance video data and explore opportunities for publishing it for analysis purpose.

In this work we recognized that a large number of video applications use blob detection and tracking as basic building blocks. For instance, the most important and widely researched issue in video surveillance research is event detection, which extensively involves blob detection and tracking. Similarly, other applications such as policy making, behavior analysis, crowd analysis, people counting also mainly depend on the blob detection and tracking. Therefore, we

propose to choose a transformation function that allows blob detection and tracking, but face, location, and time information is hidden. Specifically, we propose a hybrid method of privacy preservation that uses a combination of blurring and quantization on the whole image. The experimental results on staged data as well as 24 hours of real surveillance video show that this method produces video data which minimizes both the privacy loss and the utility loss.

3.4.1 Problem Formulation

Our main goal is to release video data such that anyone using the data cannot get any additional information about the individuals in the video. This can be achieved by hiding the identity and/or the sensitive information from the video. We have two conflicting demands here: *Utility* and *Privacy*. If we increase the privacy, the data may lose its usefulness for the intended application. On the other hand, the increase in the utility may result in disclosure of the private information. To show the trade-off between these demands, we measure the loss in privacy due to public access of the video data and the loss in the utility due to transformation. For the sake of clarity, we first provide the definition of the utility loss.

Definition 12 *Utility loss of the published video data refers to the decrease in the degree of accuracy by which analysis tasks can be accomplished with respect to the original data.*

With this definition, we formulate the research problem of video data publication as follows:

- Let V be the original video data, V' the transformed video data, and k tasks (τ_1, τ_2, ..., τ_k) are to be performed on the transformed data.
- The *Utility loss* of the published data V is denoted by $0 \leq U(V) \leq 1$, which is computed by aggregating the Utility loss values $U_j(V)$, $1 \leq j \leq k$, for the k tasks along with their associated weights that are determined based on the application requirement. Note that $U_j = 0$ implies zero utility loss i.e. the j^{th} task can be accomplished with the same accuracy as with original data, whereas $U_j = 1$ means full utility loss i.e. the j^{th} task cannot be accomplished with the transformed data.
- Let $\mathcal{F}$ be a transformation function that converts original video data V to the published video data V'.

Our goal is to find $\mathcal{F}$ that minimizes the following energy function E:

$$E = \eta\Gamma(\mathcal{F}(V)) + (1-\eta)U(\mathcal{F}(V)) \tag{3.18}$$

where η is the importance factor to have a tradeoff between the utility and the privacy. Its value can be low if the utility is more important to us and high if privacy restrictions are more severe.

3.4.2 Utility Loss Computation

Utility loss is closely associated with the application, which can be abstracted as a set of tasks. Different applications may have different sets of tasks. For example, in a video surveillance application there are four major tasks: i) to detect and recognize faces for identifying people (*who* information), ii) to detect and track blobs for finding out their activities (*what* information), iii) to detect text captions and landmarks for identifying the location (*where* information), and iv) to compute overall brightness level for identifying the time of day (*when* information).

In the statistical data publication, the traditional way to control the privacy loss is generalization [FWCY10]. The quasi identifiers, such as gender and age which cause identity leakage implicitly, are divided into groups according to the generalization algorithm and one value is chosen to represent the whole group, referred to as the generalized value. However, after performing the generalization, the data becomes less informative. The loss of information is calculated by comparing the original data with the generalized data. Using similar approach, we transform the video data using blurring, quantization, and pixelization in order to ensure the privacy. The utility loss is calculated by comparing the accuracies of the tasks before and after the transformation. The desired transformation function is the one that minimizes both the utility loss and the privacy loss.

As discussed above, the *Utility loss* $U_j(V')$ of the data V' for a task τ_j is the measure of decay in the accuracy due to data transformation. To calculate *Utility loss*, the detector (e.g. face detector, blob detector etc.) corresponding to the task (face detection, blob detection) is run over the whole video data before and after transformation. $U_j(V')$ is then calculated by considering the accuracy of the detection task after transformation ($Acc_j(V')$) and with original

data ($Acc_j(V)$). This is formulated as following:

$$U_j = 1 - \frac{Acc_j(V')}{Acc_j(V)} \tag{3.19}$$

where $Acc_j(V)$ is the accuracy of the task on data V and it is calculated as:

$$Acc = \frac{TP}{(TP + FP + FN)} \tag{3.20}$$

The true positives (TP), false positives (FP), and false negatives (FN) are defined as follows - TP: The detection corresponds to ground truth, FP: The detection does not have any corresponding ground truth, and FN: No corresponding detection for the ground truth.

Here we have omitted true negatives from calculation because the sample space is not limited i.e. the number of detections is not bounded for a frame. The overall loss in the utility U is computed by aggregating the utilities of the data for all the tasks. Precisely,

$$U(V') = \sum_{j=1}^{k} \alpha_j \times U_j(V') \tag{3.21}$$

where α_j is the normalized weight of the j^{th} task in the overall utility loss of the data. The higher the weight, the more the utility loss of the data. These weights are taken based on the goal for which the publishable data has been prepared. For example, we may want to publish the video data for the application which mainly needs blob detection. In this case, in order to compute the overall utility loss, the weight for the face detection task would be low while it would be very high for the blob detection task. For instance, in a surveillance scenario where activity detection is the prime motive, blob detection is more important than face detection, hence, it has more weight in computing the utility loss. Having said that, the weights should be determined based on application scenario and would vary significantly over multiple applications.

3.4.3 Data Transformation

Here we discuss the data transformation functions (See $\mathcal{F}$ in Section 3.4.1) that can be used to prepare the publishable data V' from the original video data V. Ideally, the function $\mathcal{F}$ should

contribute minimum to the utility loss with full privacy preservation. With this requirement in mind, we examine selective and global transformations. First we describe the limitations of selective obfuscation in Section 3.4.3. Next, in Section 3.4.3, we discuss pros and cons of various global transformation options: resolution variation (pixelization), blurring, and quantization. We find that these global transformations do not provide the best privacy-utility tradeoff when used alone. We finally propose a hybrid global transformation method in Section 3.4.3.

Selective Obfuscation

This is the most popular form of transformation explored by researchers. Computer vision techniques are used to determine the ROI in the image that are later obfuscated to hide the evidence information. This type of transformation heavily depends on the accuracy of vision algorithms. Furthermore, the large number of false detections introduce random patches in the image, limiting its utility. An accurate model of background is necessary for many multimedia tasks like activity analysis, tracking; however, it is difficult to determine whether the patchy region belongs to foreground or background. The background model needs to be updated according to the operations used to obfuscate the regions. Unlike other works, we do not make any assumption of the availability of accurate detectors, as one of our motivations for data publishing is unavailability of accurate detectors. The published video data is to be used by researchers to improve the performance of the detectors.

Global Transformation

In the global transformation, the obfuscation operation is applied on the whole image. This method of data transformation is robust as it does not depend on the detectors. At the same time, this method is too pessimistic. We transform the whole image just to hide a small portion of it. Fortunately, the background model based object detection and tracking methods work better on the globally transformed data rather than selectively obfuscated video. This is because it is easy to maintain and update the dynamic background model when the data is globally transformed. In order for the dynamic background modeling to work on selectively transformed video, we need to maintain the background models for all transformations and keep updating them. Further, the selected background model may change pixel to pixel. Therefore,

in this work we chose to globally transform the video. There are many operations that can be used to globally transform the data. Below we explore three commonly used operations [CNIB08] and discuss their limitations.

Pixelization/Resolution Variation

Pixelization is the process of effectively reducing the resolution of the image. It is done by replacing a square block by its average. If δ is the block size and $f(x, y)$ is the original image, the pixelized version $f_p(x, y)$ of the image can be obtained as follows:

$$f_p(x,y) = \frac{1}{\delta^2} \sum_{i=1}^{\delta} \sum_{j=1}^{\delta} f\left(\left\lfloor \frac{x}{\delta} \right\rfloor \delta + i, \left\lfloor \frac{y}{\delta} \right\rfloor \delta + j \right) \tag{3.22}$$

Changing the resolution of the image from high to low can be used to hide the evidence information from the video data. The reduction in resolution depends on the size of region of interest, and needs to be estimated for the worst case scenario. Decreasing the resolution of the image makes it difficult to identify the people in the video; nevertheless, the loss in the utility of the data can be drastic. The information loss is much more prominent with a slight change in the resolution of the image. Furthermore, the image loses the precious spatial information. For instance, Figure 3.2(a) shows the 47% sub-sampled version of a 320×240 pixel image; while the face detector fails at this resolution, the persons can be still identified by human beings.

Blurring

In this method we use a low pass filter (also known as Gaussian filtering) on the entire image to hide the evidence information. The Gaussian blurred image $f_b(x, y)$ is obtained by convolving the original image with a Gaussian function $G(x, y)$ i.e.

$$f_b(x,y) = f(x,y) * G(x,y) \tag{3.23}$$

with $G(x, y)$ given by

$$G(x,y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \tag{3.24}$$

where σ is the standard deviation of the Gaussian. From the privacy perspective, it can

effectively hide the evidence information present in the video; however, there are limitations of this method. Firstly, to effectively hide the evidence information, we need an estimate of the area occupied by region of interest; e.g. a bigger face area would require more blurring. Secondly, image enhancement techniques can be used to approximate the original image. If less blurring is done, the viewer can easily identify the person by looking at the image (Figure 3.2(b)) on the other hand if we blur the image excessively, the boundary of the foreground and background gets smoothed which adversely affects the detection tasks such as blob detection. Nevertheless, we can see from Figure 3.2(c) that it is very difficult to detect text in this image, both by detectors and human beings. So we conclude that blurring can help in removing high frequency evidence information (like textual information), yet it is not sufficient for effective privacy preservation.

Quantization

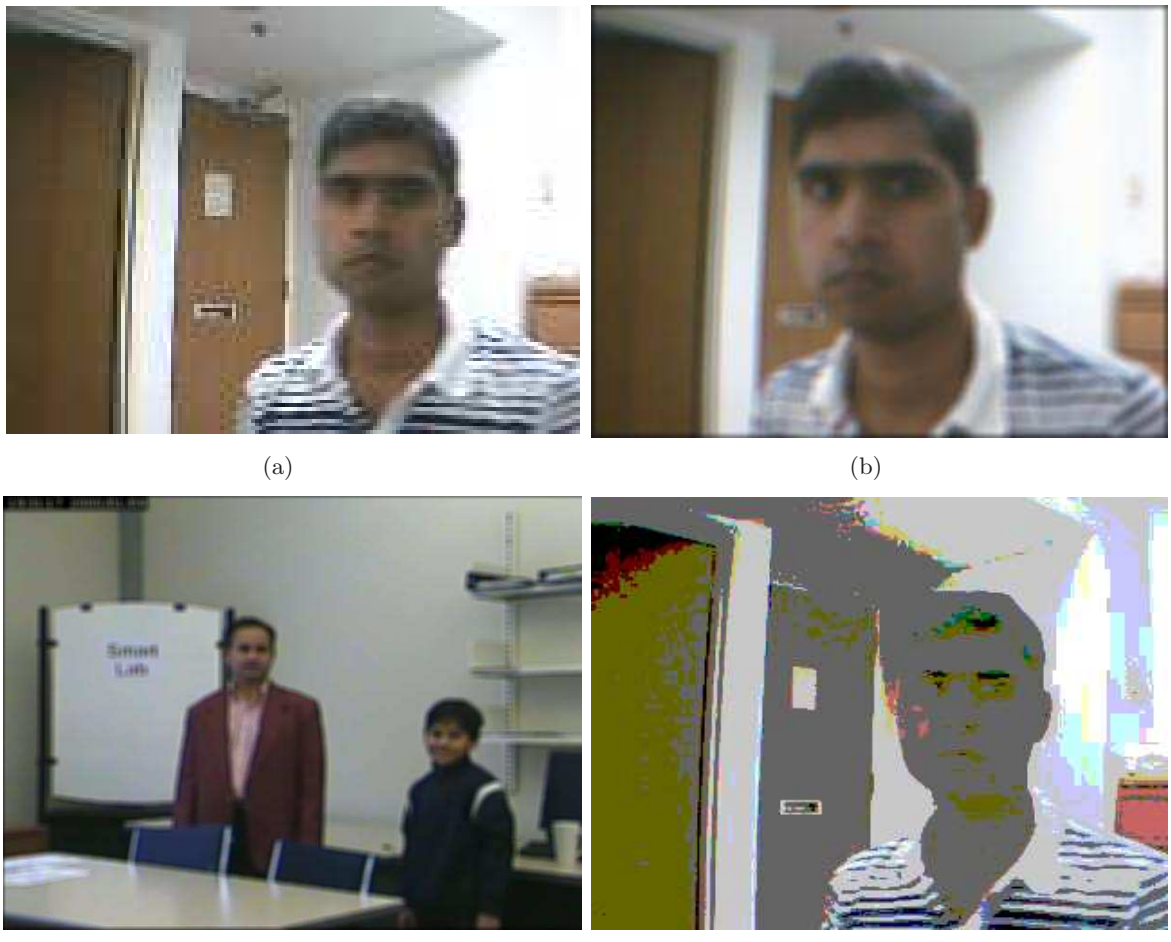
In statistical data publication, the quasi-identifiers are generally numerical values that are generalized to a representative; e.g. the age of people between 25 and 35 can be generalized to 30. Mathematically this is equivalent to the quantization of the age values. Following a similar approach, quantization can be used as a generalization tool for video (image) data. The quantized image $f_q(x, y)$ can be obtained as follows:

$$f_q = \left\lfloor \frac{f(x,y)}{q} \right\rfloor q + \frac{q}{2} \tag{3.25}$$

Image quantization introduces permanent loss of the information. In the quantization process, the pixel values are rounded to the nearest quantization level, and there is no means to recover the original value post quantization. Also, in Figure 3.2(d) we can see that it is quite difficult to identify the person from a coarsely quantized image. While the quantization is effective in hiding facial information, it performs very poorly in hiding textual information as texts are generally recognized even in a binarized image.

Proposed Hybrid Approach (Blurring and Quantization)

It can be seen that no method, alone, is sufficient to remove the privacy information effectively. We propose a hybrid method combining both “blurring” and “quantization” as it can provide



(a) (b)

(c) (d)

Figure 3.2: (a-b) Even when the face detector fails, the person can be identified by looking at the blurred image, (c) The resolution is reduced to 47% for the face detector to fail, still face can be identified, (d) It is difficult to identify the person from a coarsely quantized image.

better utility for a given privacy than individual operations of blurring and quantization. In the hybrid approach, we can apply blurring first or quantization first. In the first case, we analyze the video to determine how much blurring is needed to suppress the high frequency evidence information; based on which we apply a Gaussian filter of appropriate size. The blurred image is then coarsely quantized to hide the identity i.e.

$$f_{bq} = \left\lfloor \frac{f_b(x,y)}{q} \right\rfloor q + \frac{q}{2} \tag{3.26}$$

where f_b is given in Equation 3.23 and f_{bq} is the first blurred and then quantized image. Effectively, we first remove the high frequency private information from the image via blurring and then introduce random non-recoverable high frequency noise via quantization to hide the identity information. In the second case we reverse the order of the transformations and apply quantization first followed by blurring i.e.

$$f_{qb}(x,y) = f_q(x,y) * G(x,y) \tag{3.27}$$

where f_q is given by the Equation 3.25 and f_{qb} is the first quantized and then blurred image.

We found in the experiments that the proposed method transforms the video to a form where human beings cannot easily identify the people in the video; however, some application tasks like blob detection and tracking can still be accomplished. It is shown that the proposed hybrid method systematically removes evidence information while preserving the desired utility of the data. Particularly, the first blurring and then quantization approach is able to obtain the best tradeoff between the privacy and the utility.

While the superiority of the proposed method is validated by conducting extensive set of experiments on large amount of the video, the privacy loss in those experiments is still calculated manually. It is hard to calculate the privacy loss when the length of the video is large. We make the following two observations to cope with this problem:

- In a surveillance video, the background regions that cause evidence detection are of the same size throughout the video. Therefore, the parameters determined for a fixed length of the video are applicable for the whole video irrespective of its length.
- The transformation required to hide the facial information depends on the face size and

its proportional to the size of facial region. Therefore, the part of the video that contains biggest face size will determine the transformation parameters required to hide faces in the whole video.

Hence, we can address the scalability problem if we are able to find the portion of the video that has the largest face size. Fortunately, the surveillance cameras are generally fixed and the minimum person-to-camera distance can be easily determined. This domain knowledge can be exploited to find a representative clip from the video and the proposed method can be applied on that small video clip to obtain the transformation parameters. Now, any video clip from this camera can be transformed with these parameters, solving the problem of scalability. In summary, following are the steps that can be taken to publish surveillance video recorded from a given camera:

1. Analyze the video from the camera and select a segment of the video in which people are walking closest possible to the camera and frontal faces are visible. This is called representative video. The representative video can also be recorded with help of actors by making them walk as close to the surveillance camera as possible with frontal face visible.

2. Obtain the transformation parameters for this small segment of the video according to the proposed method of first blurring then quantization. Note that the privacy loss calculations are done on the representative video, while the utility loss calculations can be done for the whole video as it is done automatically.

3. Now any clip that is recorded from the same surveillance camera can be transformed using those parameters.

In this way, after initial preprocessing step, the method can be applied online to any length of the video.

3.4.4 Experiments and Results

The application scenario considered for experiments consists of two tasks: blob detection and tracking. Both the tasks are considered equally important, so the associated weights for utility calculation are $\alpha_1 = 0.5$ and $\alpha_2 = 0.5$. For blob detection, a Gaussian mixture models (GMM)

Table 3.2: Description of the video data used in experiments

Data	Type	Total frames	Activity frames	People (ρ)	Resolution	Duration	Scenario
Video1	Mock	6650	5045	2	320×240	30 Minutes	Indoor
Video2	Mock	3940	2340	4	320×240	30 Minutes	Outdoor
Video3	**Real**	**28216**	**17218**	20	704×480	**24 Hours**	Indoor

based adaptive background model [SG99] is used; and the blob detections in the original and transformed data are compared based on the blob location and area to detect TP, FP, and FN. The most important difference between the tracking and the blob detection is that the blob detection works on individual frames while in the tracking we associate targets across frames [MSS08]. To calculate the tracking accuracy, we compare the blob associations in the original and the transformed video. The correct and incorrect associations are determined by comparing pixel histograms to calculate TP, FP, and FN for tracking. The utility loss of both tasks is combined using Equation 3.21 to calculate overall utility. Since the accuracies are calculated with respect to the detections in the original data, $Acc(V)$ is unity for both the tasks.

The value of η is chosen slightly more than 0.5 (i.e. 0.6) to simulate a scenario in which the privacy is more important than the utility. While the effect of the sensitive information is kept out of scope of this work and will be studied in future, in these experiments we consider an activity as the sensitive information. Since the activity information is present in all the original and transformed videos, the sensitivity index is unity for all experiments i.e. $\Psi = 1$.

The experiments have been designed to study the following issues:

- Effect of the hidden inference channels on the privacy loss from the published video data.
- Effectiveness of different transformation methods in reducing the privacy loss.
- Optimum amount of blurring and quantization required to minimize the energy function (see Equation (3.18)) for a given application.
- Effectiveness of the hybrid approach in which we use both blurring and quantization simultaneously.

Dataset

The experiments are conducted with set of three videos with different characteristics (see Table 3.2). Two videos are recorded in indoor surveillance settings and one in outdoor setting. Video1

(a) Video1 (b) Video2 (c) Video3

Figure 3.3: Representative frames from the three video clips.

is recorded in a university lab (indoor) and contains *who*, *what*, *when* (through glass window), *where* (through text legend) evidence. Similarly, Video2 is recorded at the entrance of floor where the lab is situated (outdoor) and contains *who*, *what*, *when* (sunlight), *where* (through text legend and university logo) evidence. The real surveillance video (Video3) has *who* and *what* evidences. Although in these experiments the main source of *where* evidence has been text legend, it can come from other sources as well e.g. familiar places, company logos, familiar figures etc. In order to remove the evidence all the sources which can lead to the detection of that evidence should be transformed. The representative frames from the videos are shown in Figure 3.3. Note that for the purpose of anonymity, we show only blank video frame (without any person in it) for Video 3 throughout this chapter, e.g. Figure 3.3c.

For all three videos, people performed walking activity ($n_1 = 1$, $\xi_1 = walking$). The only source of where evidence was text legend ($n_2 = 1$, $\lambda_1 = legend$), and the time was divided into two slots i.e. $n_3 = 2$, $\Delta t_1 = day$ and $\Delta t_2 = night$. With these coefficients, the group sizes for the first two videos are determined as follows: the G_{wt} is a large number (i.e. 100000) because without the knowledge of *when* and *where* information it can be assumed that lots of people do lots of activities in the world; however, if we also have the *when* evidence, the group size ($G_{wt,wn}$) reduces to approximately 10000. In the same way, by knowing the *where* information with the *what* information, the group size ($G_{wt,wr}$) reduces to a small value, e.g. 10 in our case. This number was chosen based on the information that 10 graduate students used to come to the graduate lab (we named it "Smart Lab") at our institution. Finally, with all three evidences *what*, *when* and *where*, the cluster size ($G_{wt,wn,wr}$) further reduces to a very small number. In our case, this number was 5. This is because on average only five students used to

come to the lab during day time. The real video was recorded at the main building entrance, hence the group size for evidence *what* and *where* ($G_{wt,wr}$) is 1000 and ($G_{wt,wn,wr}$) is 100. Other coefficients are the same as in the other videos.

Privacy Assessment: Implicit Vs Explicit

Since current automatic evidence detectors are still far behind a human's capabilities of evidence detection, we performed a user study to calculate the privacy loss and demonstrate the proposed framework. In this study, our three colleagues participated. The users were asked the following three questions: 1) is the face recognizable? 2) is the time when the video was shot available? and 3) is the location recognizable? These three questions were asked to conform with the I_{wo}, I_{wn} and I_{wr} values, respectively, obtained using our privacy assessment method. The overall answer is calculated as "No" only if all users answer in "No", otherwise it is considered "Yes". Note that we did not ask questions like 'Can you recognize the person?' to minimize the subjective component in evaluation. The answer to such questions heavily depends on the users' prior knowledge.

We provide the answers of the above three questions for video1 in Table 3.3 (for varying degree of blurring) and Table 3.4 (for different quantization steps). The 'Machine detected' column shows the result of automatic techniques run on the machine, while the 'Human detected' column contains the user opinion. The privacy loss is provided in last two columns, which is computed based on machine-detected and human-detected options. The boldfaced values in the table show the nonzero identity leakage that occurs due to implicit channels. In these cases, even when the face is removed (i.e. I_{wo} is zero), the privacy loss would be non zero. This situation can be observed in Table 3.4 in the rows corresponding to quantization 90 and 120. The explicit identity leakage is zero while the implicit identity leakage is still 0.2 leading to nonzero privacy loss. Therefore, focusing on only explicit identity leakage may leave loop holes in hiding identity, while the proposed method provides more robust measures of privacy.

We also observe that in many cases the privacy loss based on the machine detections predicts zero value, whereas the human detected privacy loss is still significant. This result mainly shows that the current automatic methods of evidence detection are not reliable; for instance the face could still be recognized by humans while the face detector failed. Similar results are obtained

Table 3.3: Privacy loss for video1 with different degrees of blurring.

	who		*where*		*when*	
b	M*	H*	M*	H*	M*	H*
1	Yes	Yes	Yes	Yes	No	Yes
2	Yes	Yes	No	Yes	No	Yes
4	No	Yes	No	No	No	Yes
6	No	No	No	No	No	No
8	No	No	No	No	No	No
10	No	No	No	No	No	No

*M and H represent Machine-detected and Human-detected options, respectively.

	Identity leakage				Privacy Loss	
	Implicit I_{im}		Explicit I_{ex}		Γ	
b	M*	H*	M*	H*	M*	H*
1	0.2	**0.4000**	1	1	1	1
2	0.0	**0.4000**	1	1	1	1
4	0.0	**0.0002**	0	1	0	1
6	0.0	0.0000	0	0	0	0
8	0.0	0.0000	0	0	0	0
10	0.0	0.0000	0	0	0	0

Table 3.4: Privacy loss for video1 with different quantization steps.

	who		*where*		*when*	
q	M*	H*	M*	H*	M*	H*
1	Yes	Yes	Yes	Yes	No	Yes
30	Yes	Yes	Yes	Yes	No	Yes
60	Yes	Yes	No	Yes	No	No
90	No	No	No	Yes	No	No
120	No	No	No	Yes	No	No
150	No	No	No	No	No	No

*M and H represent Machine-detected and Human-detected options, respectively.

	Identity leakage				Privacy Loss	
	Implicit I_{im}		Explicit I_{ex}		Γ	
q	M*	H*	M*	H*	M*	H*
1	0.2	**0.4**	1	1	1	1.0
30	0.2	**0.4**	1	1	1	1.0
60	0.0	**0.2**	1	1	1	1.0
90	0.0	**0.2**	0	0	0	0.2
120	0.0	**0.2**	0	0	0	0.2
150	0.0	0.0	0	0	0	0.0

for video2 (Table 3.5 and Table 3.6) and video3 (Table 3.7 and Table 3.8). Consequently, in the remaining experiments we calculate the privacy loss by considering the human detections.

Effect of Data Transformation Methods on Privacy Loss and Utility Loss

In this experiment, we explore the effect of various data transformation methods on the privacy loss and the utility loss. The following data transformation methods are used: selective obfuscation, blurring only, quantization only, hybrid (both quantization and blurring). We also study trade off between the privacy and utility losses and compute the energy function using Equation 3.18. The privacy assessment is explained in detail in the previous section.

Table 3.5: Privacy loss for video2 with different degrees of blurring.

	Implicit identity leakage I_{im}		Explicit identity leakage I_{ex}		Privacy Loss Γ	
Blurring (b)	M*	H*	M*	H*	M*	H*
1	0	0.8	1	1	1	1
2	0	0.8	1	1	1	1
4	0	0.8	0	1	0	1
6	0	0.8	0	0	0	0.8
8	0	0.8	0	0	0	0.8
10	0	0.4	0	0	0	0.4

	who		*where*		*when*	
Blurring (b)	M*	H*	M*	H*	M*	H*
1	Yes	Yes	No	Yes	No	Yes
2	Yes	Yes	No	Yes	No	Yes
4	No	Yes	No	Yes	No	Yes
6	No	No	No	Yes	No	Yes
8	No	No	No	Yes	No	Yes
10	No	No	No	Yes	No	No
12	No	No	No	No	No	No

*M and H represent Machine-detected and Human-detected options, respectively.

Table 3.6: Privacy loss for video2 with different quantization steps.

	Implicit identity leakage I_{im}		Explicit identity leakage I_{ex}		Privacy Loss Γ	
Quantization (q)	M*	H*	M*	H*	M*	H*
1	0	0.8	1	1	1	1
30	0	0.8	1	1	1	1
60	0	0.8	1	1	1	1
90	0	0.8	0	1	0	1
120	0	0.8	0	1	0	1
150	0	0.4	0	0	0	0.4

	who		*where*		*when*	
Quantization (q)	M*	H*	M*	H*	M*	H*
1	Yes	Yes	No	Yes	No	Yes
30	Yes	Yes	No	Yes	No	Yes
60	Yes	Yes	No	Yes	No	No
90	No	Yes	No	Yes	No	No
120	No	Yes	No	Yes	No	No
150	No	No	No	Yes	No	No

*M and H represent Machine-detected and Human-detected options, respectively.

Table 3.7: Privacy loss for video3 with different degrees of blurring.

	Identity leakage				Privacy Loss Γ	
	Implicit I_{im}		Explicit I_{ex}			
Blurring (b)	M*	H*	M*	H*	M*	H*
1	0	0.0002	1	1	1	1
2	0	0.0002	0	1	0	1
4	0	0.0002	0	1	0	1
6	0	0.0002	0	0	0	0.0002
8	0	0.0002	0	0	0	0.0002
10	0	0.0002	0	0	0	0.0002

	who		*where*		*when*	
Blurring (b)	M*	H*	M*	H*	M*	H*
1	Yes	Yes	No	No	No	No
2	No	Yes	No	No	No	No
4	No	Yes	No	No	No	No
6	No	0	No	No	No	No
8	No	0	No	No	No	No
10	No	No	No	No	No	No

*M and H represent Machine-detected and Human-detected options, respectively.

Table 3.8: Privacy loss for video3 with different quantization steps.

	Identity leakage				Privacy Loss	
	Implicit I_{im}		Explicit I_{ex}		Γ	
Quantization (q)	M*	H*	M*	H*	M*	H*
1	0	0.0002	1	1	1	1
30	0	0.0002	0	1	0	1
60	0	0.0002	0	1	0	1
90	0	0.0002	0	0	0	0.0002
120	0	0.0002	0	0	0	0.0002
150	0	0.0002	0	0	0	0.0002

	who		*where*		*when*	
Quantization (q)	M*	H*	M*	H*	M*	H*
1	Yes	Yes	No	Yes	No	Yes
30	No	Yes	No	Yes	No	Yes
60	No	Yes	No	Yes	No	No
90	No	Yes	No	Yes	No	No
120	No	Yes	No	Yes	No	No
150	No	No	No	Yes	No	No
150	No	No	No	No	No	No

*M and H represent Machine-detected and Human-detected options, respectively.

Detect and Hide Evidence Regions

In this method the evidence regions of the image are selected using computer vision techniques and consequently obfuscated. However, since these techniques are usually not fully accurate, there may be cases where the evidence regions in a few video frames are left undetected and remain visible. In Figure 3.4, the first row shows the results of face detection, and the second row the results of a method based on hiding the facial information. In the first frame (Figure 3.4a), the face is hidden properly, which helped in accurately removing the facial information from the image. However, in Figure 3.4b, the face region is incorrectly detected and therefore incorrectly obscured, while in Figure 3.4c, the face is not detected at all and remains visible. Note that if the face is seen in even one frame, the identity is revealed.

Even with error-free accurate face detectors, hiding faces is not enough for preserving the privacy. The implicit inference channels due to evidences *when* and *where* also need to be blocked; which again needs accurate detectors such as a text detector, landmark detector, etc. We applied a text detector on the video clip and found that it fails to detect text even when the writing is clearly visible and readable by a human. One of such instances can be seen in Figure 3.4c. This shows that we need a data transformation method that either does not depend on the accuracy of the vision algorithms or provides robust the privacy preservation with the existing algorithms.

The utility of the data is also affected adversely if there are false detections. If the obfuscated region does not belong to the actual foreground (for example, false face detections), the obscured

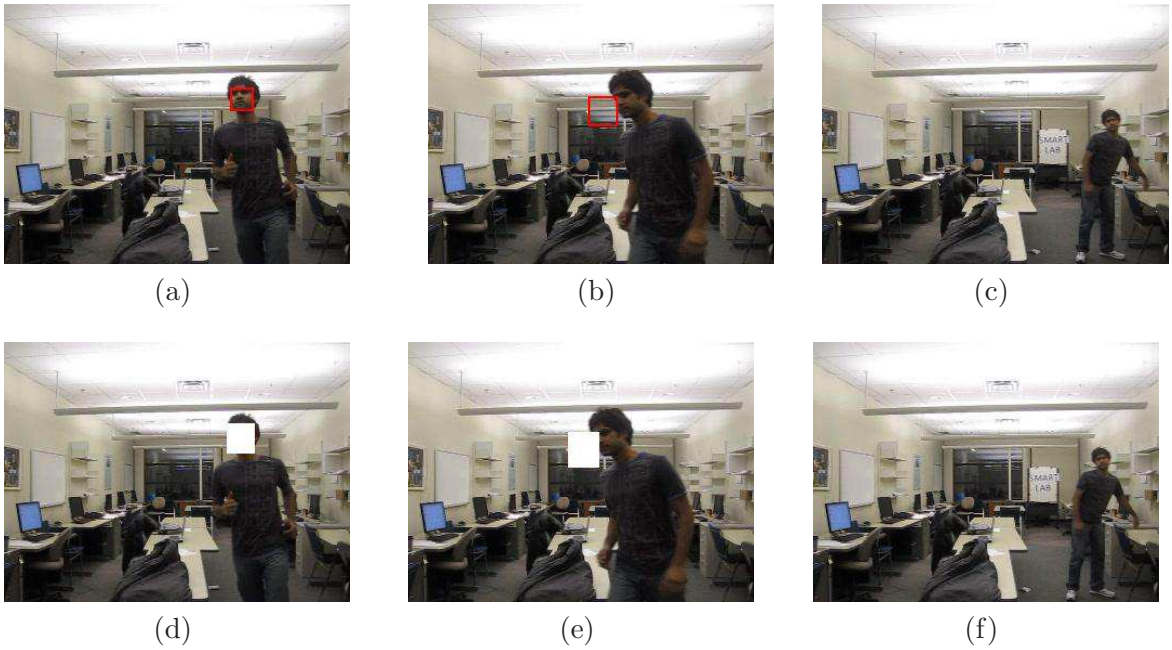


Figure 3.4: Row 1 shows the results of the face detection and row two the transformed data.

Table 3.9: Privacy loss, utility loss, and Energy calculation for selective obfuscation method.

	Video1			Video2			Video3		
Operation	Γ	U	E	Γ	U	E	Γ	U	E
Quantization	0.999	0.108	0.643	1.000	0.222	0.689	0.974	0.183	0.658
Blurring	0.999	0.188	0.675	1.000	0.248	0.699	0.974	0.214	0.670

region will produce a false positive in the blob detection. We applied a face detector on the transformed video and calculated the utility loss. The resulting privacy and utility loss are shown in Table 3.9. We observe a large values of energy function which implies that this method is not effective in providing the privacy and the utility simultaneously.

Blurring

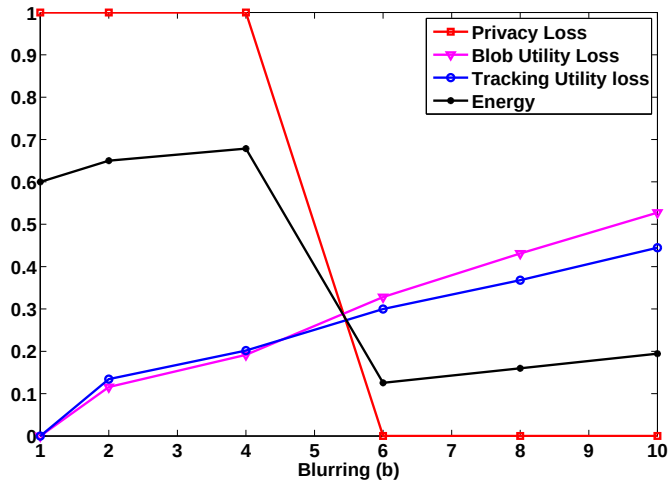
In this experiment the images are uniformly blurred to hide the private information. In this way, this method does not require ROI detection algorithms and still provides better privacy preservation, blocking both implicit and explicit channels of the identity leakage. Here, a Gaussian low pass filter is applied on the whole image. The degree of blurring is defined as the size of the filter used for blurring; a blurring of degree b means a $b \times b$ Gaussian filter is used to blur the images (with $\sigma = b$). The larger the value of b, the greater the image is blurred. The resulting privacy loss, utility loss, and energy function are shown in Figure 3.5. Initially

the privacy loss is maximum (leading to $E = 0.6$) when no transformation is done. The utility loss of blob detection and tracking tasks increases as we increase the degree of transformation causing energy to monotonically increase, until the transformation is sufficient to hide one or more evidences. The energy value encounters a dip whenever an evidence changes from detected to not detected, for instance the energy value at $b = 6$ for video2 in Figure 3.5. However, the desired minimum value of the energy is generally achieved when all the evidences are removed (except *what* which cannot be removed). For Video1 and Video3, the optimum values of the energy function are obtained for $b = 6$. In Video2, the source of *where* evidence (text legend and university logo) was close to the camera and occupied a relatively larger portion of the image. Also, people walked quite close to the camera, so we had to apply a large amount of blurring ($b = 10$) in order to remove the privacy information. The minimum value of the energy is still around 0.53, which is more compared to Video1 (0.13) and Video3 (0.077).

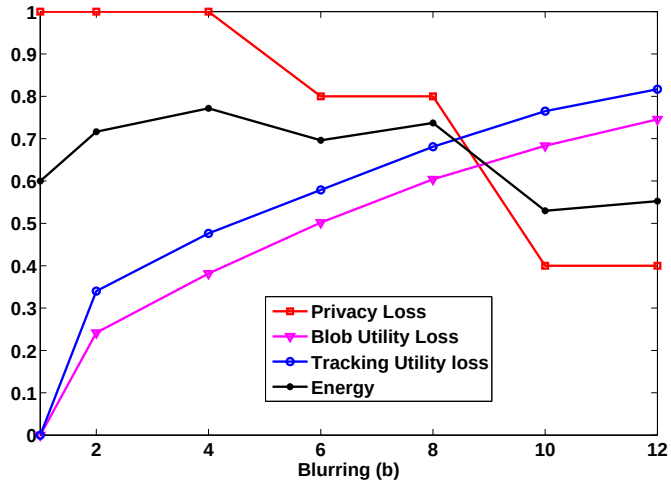
Figure 3.6 shows the results of this approach with the corresponding blurring values. In this figure, it can be easily observed that blurring removes the private information effectively; however, the problem with this method is that it is generally too pessimistic and destroys image quality which reduces the utility of the data. Therefore, we next examine the quantization method.

Quantization

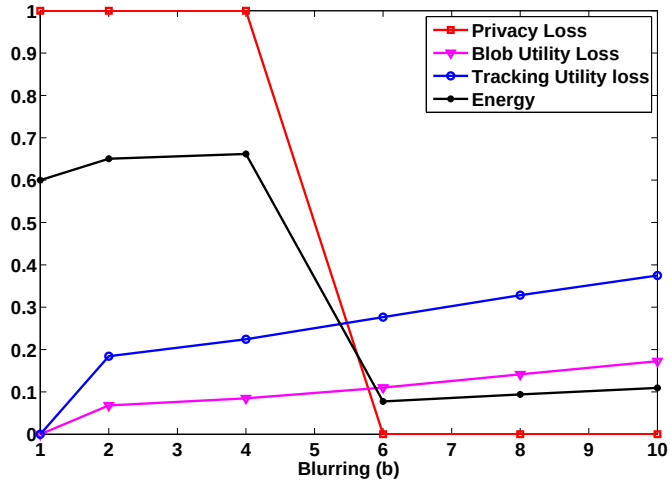
In this experiment the images are quantized to hide the private information. The pixel values are quantized with the varying quantization step q using equation 3.25. The resulting privacy loss, utility loss, and energy function are shown in Figure 3.7. For Video2 and Video3, the value of the energy function are slightly higher than what we got with blurring (0.5419 and 0.1952 respectively), while for Video1 it is lower (0.111). The higher energy values for Video2 and Video3 are due to relatively dynamic nature of the background in Video2 (moving trees and varying sunlight) and Video3 (moving escalators, light and tables, chairs etc). Also, in contrast to Video2 and Video3, Video1 is a low resolution and the text is very small, so small value of quantization is enough to remove the privacy information and we get a better energy value. Figure 3.8 shows sample results of blob detection for three different videos with corresponding quantization values which provide minimum energy. Similar to blurring, this method



video1



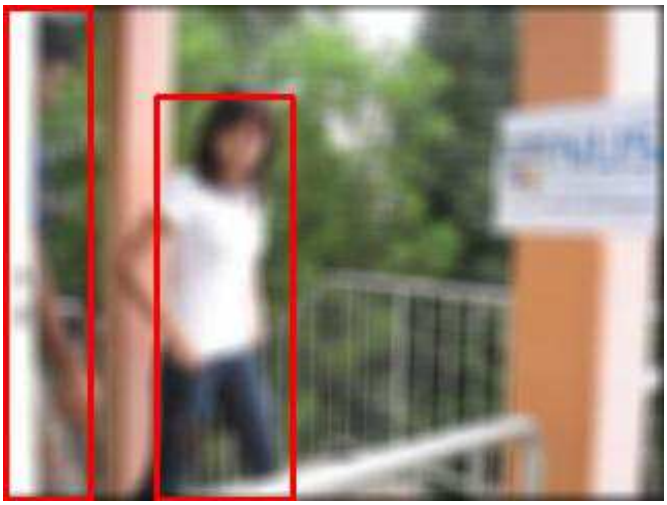
video2



video3

Figure 3.5: Privacy loss, utility loss, and energy with different degrees of blurring.

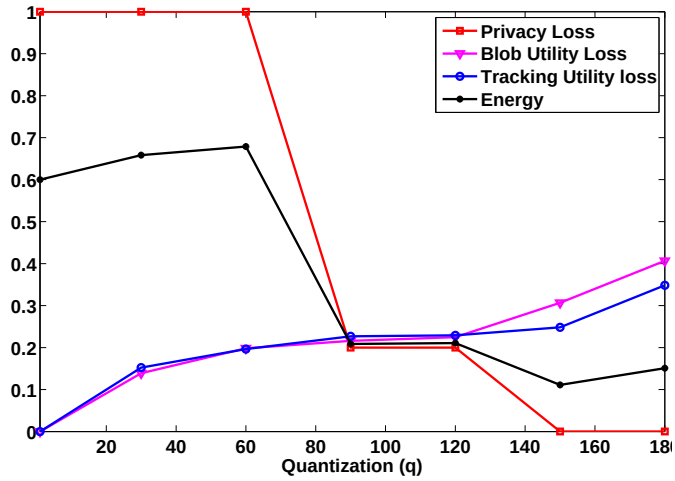
video1, $b = 6$



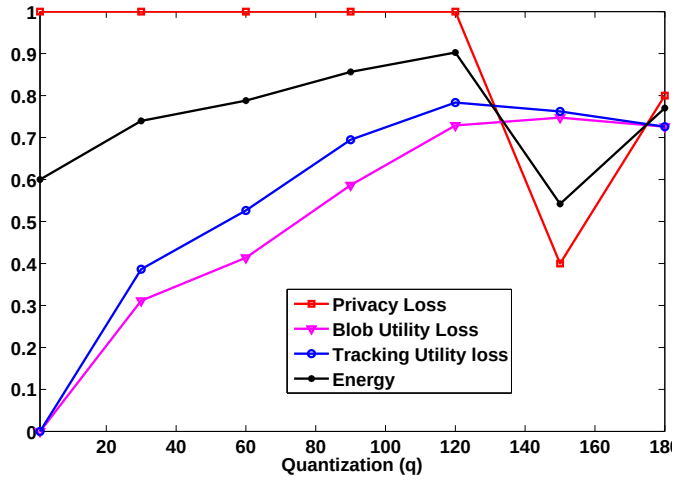
video2, $b = 10$

video3, $b = 6$

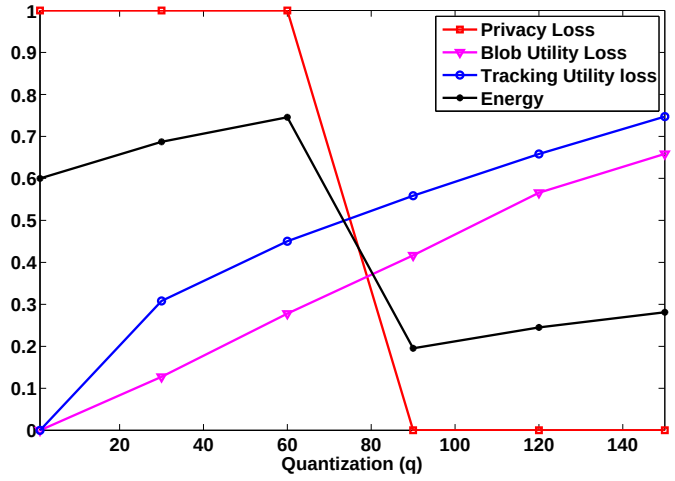
Figure 3.6: The images are blurred to hide the identity information.



video1

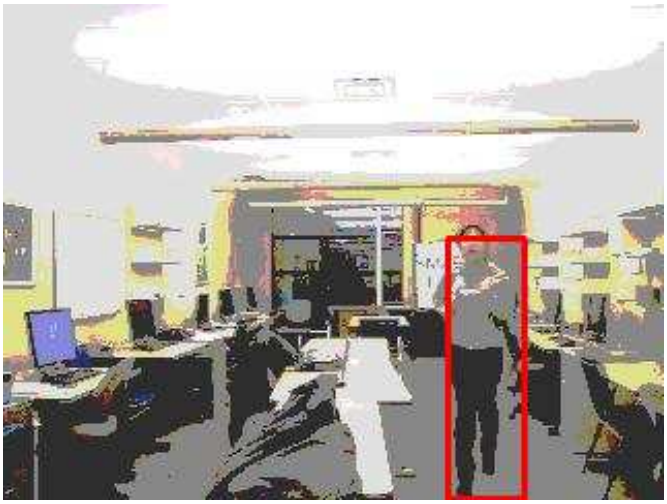


video2

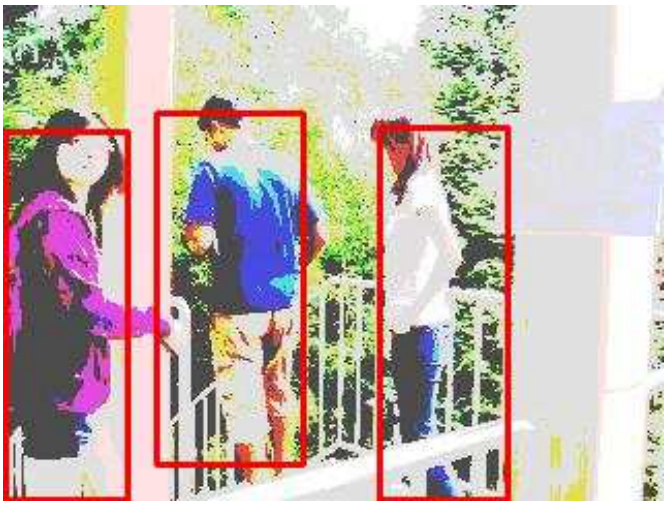


video3

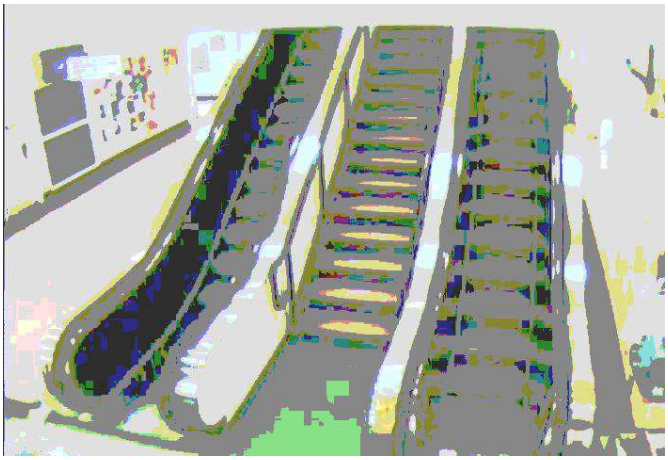
Figure 3.7: Privacy loss, utility loss, and energy with different degrees of quantization.



video1 ,$q = 90$



video2 ,$q = 150$



video3 ,$q = 150$

Figure 3.8: The images are quantized to hide the identity information.

is also effective in hiding privacy information; however, it failed to show any improvements over blurring.

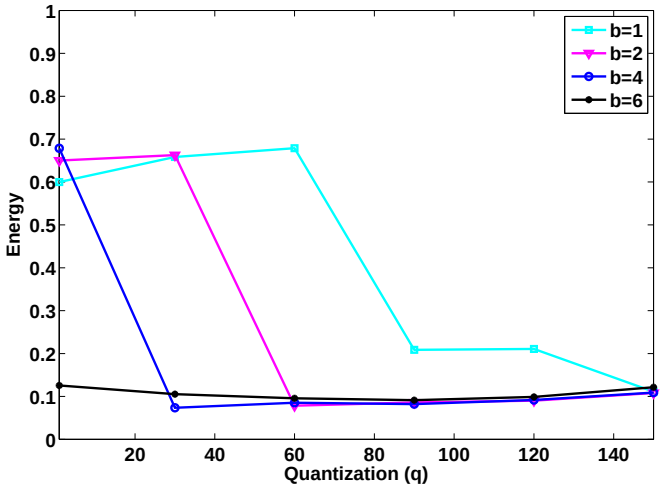
Hybrid Approach

The blurring operation removes the high frequency evidence information whereas the quantization is effective in destroying the low frequency evidence information. Both blurring and quantization, when used alone, lead to disadvantages. This observation lead us to a better hybrid transformation function, which involves a combination of blurring and quantization. In this case, the obvious questions are: how much blurring and how much quantization should be done, whether blurring should be done first and then quantization should be done, or quantization should be followed by blurring. To resolve these questions, we performed the following experiments:

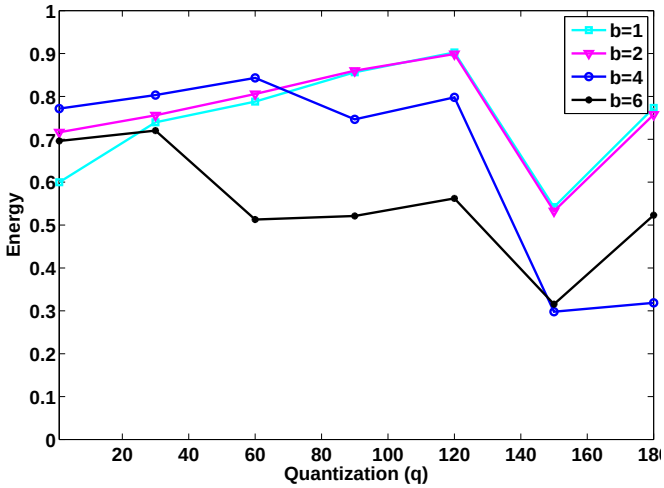
- Case #1 *First Blurring Then Quantization*: In this experiment, we fix the degree of blurring and vary the quantization step. The resulting values are shown in Figure 3.9.

- Case #2 *First Quantization Then Blurring*: Here we fix the quantization step and vary the degree of blurring. Figure 3.11 shows the resulting values.

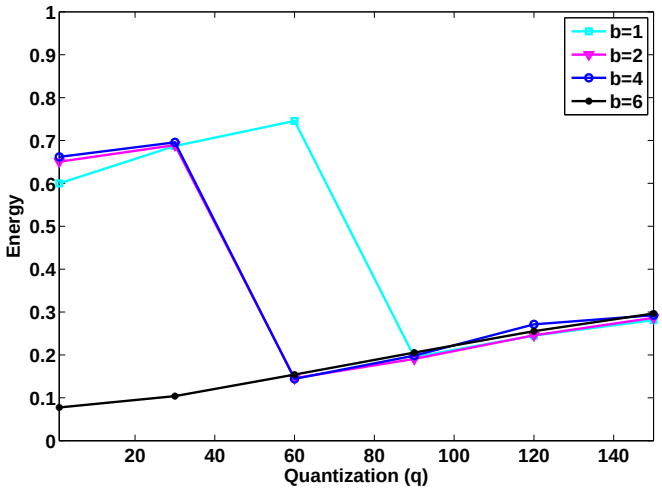
It is observed that both the methods are effective in hiding private information. However, we can notice in Figure 3.9 and Figure 3.11 that first blurring and then quantization approach is able to achieve minimum values of energy function (Equation 3.18) for all three videos. For Video1 the minimum energy is 0.073 which is obtained for $b = 4$ and $q = 30$. Similarly, the minimum energies for Video2 and Video3 are 0.297 and 0.077 which are obtained at $b = 4, q = 150$ and $b = 6, q = 1$ respectively. Figure 3.10 and Figure 3.12 show the corresponding best results for the two hybrid approaches. First blurring then quantization approach works better because blurring first removes the high frequency information and quantization adds the white noise with flat spectrum which uniformly distorts the image. In the case of first quantization and then blurring, the white noise is added first and blurring removes both the high frequency noise and the upper band of white noise added to distort the image. This uneven addition of the quantization noise negatively affects the utility of the image.



video1



video2



video3

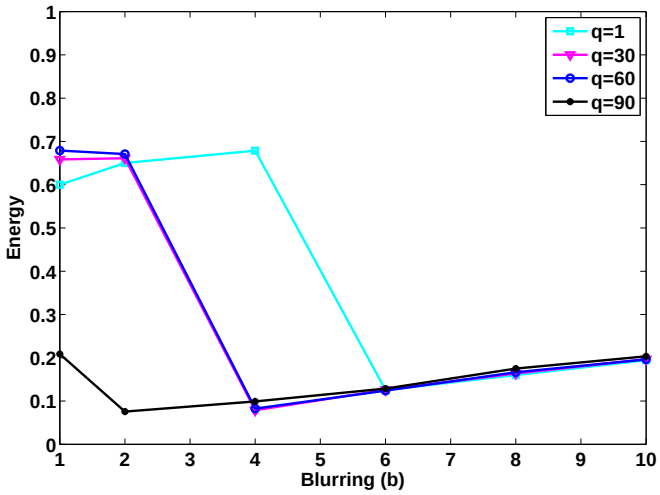
Figure 3.9: Privacy loss, utility loss, and energy with first blurring then quantization hybrid approach.

video1
$b = 4$, $q = 30$

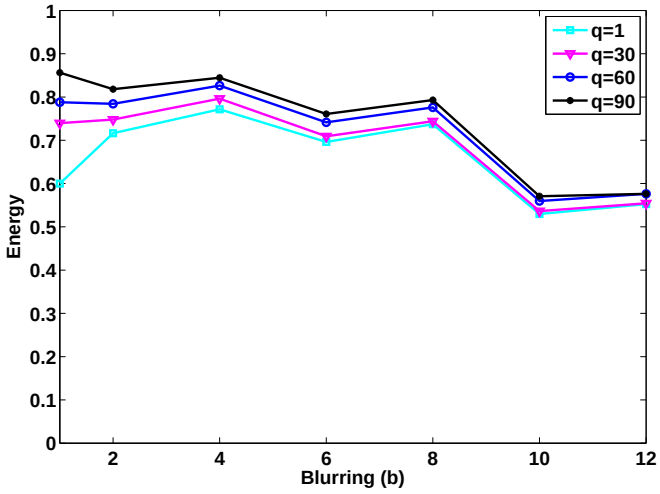
video2
$b = 4$, $q = 150$

video3
$b = 6$, $q = 1$

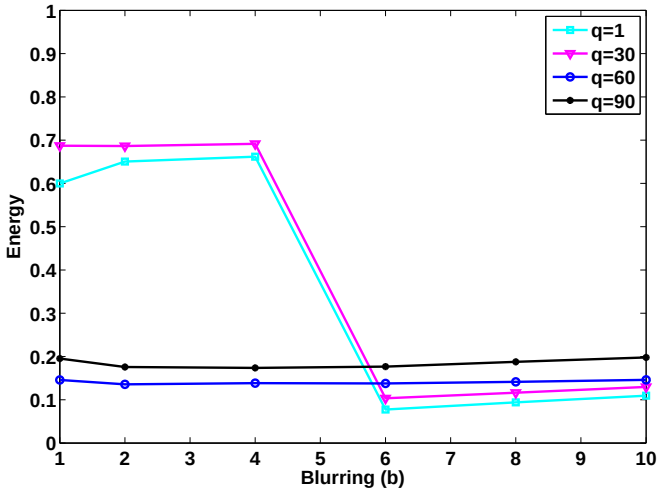
Figure 3.10: The resultant images when first blurred and then quantized.



video1



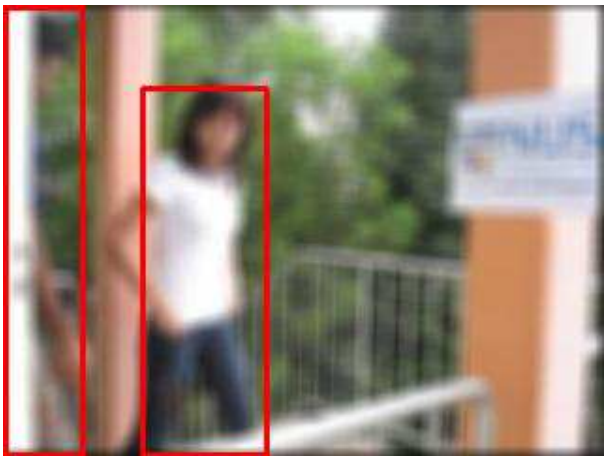
video2



video3

Figure 3.11: Privacy loss, utility loss, and energy with first quantization and then varying degrees of blurring.

video1
$q = 90$, $b = 2$



video2
$q = 1$, $b = 10$

video3
$q = 1$, $b = 6$

Figure 3.12: The resultant images when first quantized and then blurred.

3.4.5 Discussion

Experiments show that in some scenarios the implicit channels can cause severe privacy loss, hence these channels should be blocked before the data is published. The minimum energy value obtained for selective obfuscation method is 0.643 which is quite high compared to other methods which are around 0.1. This happens because the privacy loss is always close to unity in the selective obfuscation method as the detection methods are not totally reliable. Hence selective obfuscation is often not the best choice for privacy preservation. In the proposed method, between blurring and quantization methods of data transformation, blurring generally results in lower energy values hence it is considered superior than the quantization. However, the best energy values are obtained by using the hybrid approach i.e. first blurring by small amounts and then quantizing as required.

3.5 Summary & Conclusions

A robust privacy model is paramount for privacy-aware video surveillance. Current privacy models have only focused on the explicit channels of identity leakage (e.g. facial information) while other implicit channels (*what, when* and *where*) are generally ignored. In this chapter we propose a model that considers both implicit and explicit channels of identity leakage. The privacy loss is modeled as a product of the identity leakage and the presence of sensitive information; measured as sensitivity index. The model has been applied for privacy-aware publishing of surveillance video data. Maintaining a tradeoff between two conflicting demands, the privacy and the utility, of the video data is crucial for any application. In this chapter we have proposed computational model of the utility loss for the applications of data publication. It is found that a hybrid global transformation approach of blurring and quantization best serves the purpose of transforming video data. In summary, following are specific conclusions:

- The implicit channels can cause significant privacy loss even when the facial information is not present. Therefore, blocking implicit channels is also equally important.
- Detect and hide approach is not reliable and provides a bad tradeoff between the privacy and the utility.
- Hybrid approach provides better tradeoff.

- First blurring then quantization provides better tradeoff.

The proposed privacy model is for a single camera video. In the case of multiple camera videos, there will be additional identity leakage channels. In the next chapter, we enhance the proposed model in order to measure privacy loss for a multi-camera scenario.

Chapter 4

Enhanced Privacy Model for Multi-Camera Video

The implicit channels of identity leakage play a very important role in determining the privacy loss that could occur from a video. In the previous chapter, we found that removing the facial information is necessary for the privacy preservation, but it is not sufficient and the privacy loss could still occur through the implicit channels of identity leakage. However, in that work we only considered the privacy loss from a single camera video. The access to multiple camera videos may cause additional privacy loss in the following ways:

- The adversary can correlate persons in multiple video streams and observe more activities resulting in increased chances of the identity leakage and the privacy loss. For example, from a single camera generally we cannot infer what places a person visits or whom he meets over a period of time at different places. But this information can be easily extracted by correlating people over multiple camera videos.
- Imagine the adversary identifies one person in a camera video. S/he can use this information to infer the identity of other people in the video. Further, this leakage can be propagated to other camera videos if the adversary is able to correlate people across multiple videos.

In the previous model, the evidence information is measured as a binary variable for the whole video clip. For a single camera video it is sufficient because the identity leakage through

event patterns is not significant. However, in case of multi-camera video, the events may be spatio-temporally distributed; and as discussed above, the event patterns can significantly contribute to the identity leakage. Therefore, in order to measure the identity leakage from the multi-camera video, we need to analyze evidences at a higher granularity of events.

In this chapter we emphasize the implicit inference channels in the case of multi-camera surveillance videos and provide an enhanced model for the privacy loss assessment. In the proposed model, we first divide the video into a sequence of events and then represent these events using propositional logic statements. Information extraction is applied to these events to find different types of evidences *what, when* and *where* (we assume that the *who* evidence has already been blocked in the video). The evidence information is used to measure the identity leakage that can occur due to each event. To incorporate the identity leakage due to event pattern knowledge of the adversary, we prepare event lists for each target and fuse the identity leakages from all the events in the list using the anonymity based approach. Finally, the privacy loss is modeled as a product of the identity leakage and the presence of sensitive information.

While the proposed event based privacy model incorporates the breach in the current privacy methods, it also integrates seamlessly with currently growing research on event-based semantic representation of the video [DvGN+08], [PF11], [FBRG11]. The model measures the privacy at the semantic level in comparison to the earlier approaches that are mainly based on region of interest (RoI) like face and blob.

The main contributions of this chapter are summarized as:

- An enhanced model for determining the identity leakage which incorporates the events in the video.
- An enhanced model that calculates the identity leakage for each individual in the video.
- The adversary's knowledge is precisely stated and its effect on the identity leakage is calculated.

The chapter is laid out as follows. The identity leakage is modeled in Section 4.1 followed by the privacy loss assessment in Section 4.2. We provide experimental results in Section 4.3 and finally conclude the chapter with a discussion on deployment of method in real systems and its limitations in Section 4.4. Finally we conclude chapter in Section 4.5.

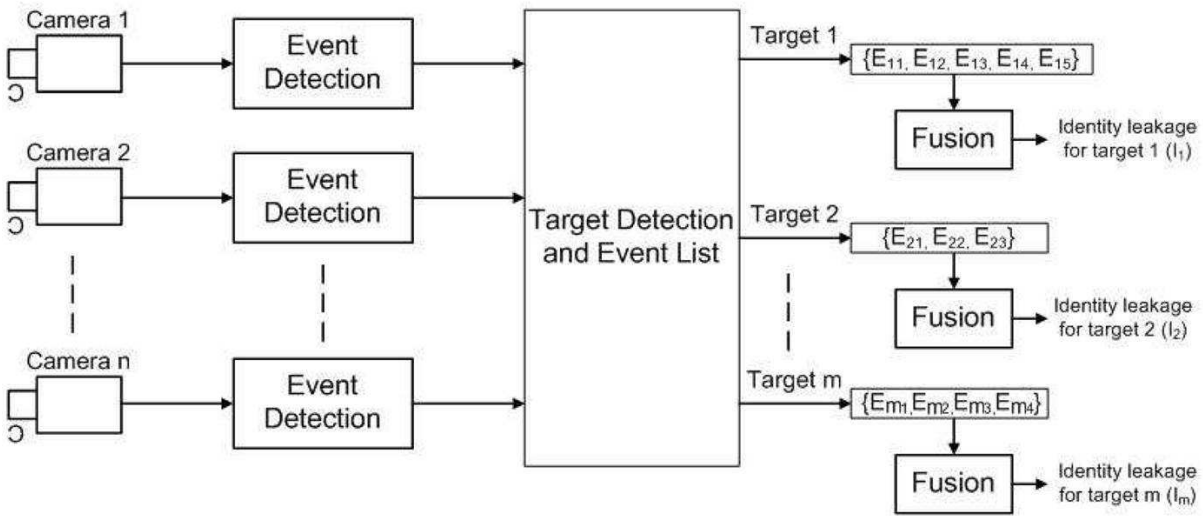


Figure 4.1: The framework for **Identity Leakage Analysis**. In this Figure targets are used to denote individuals in the video.

4.1 Identity Leakage

The aim of an adversary is to establish a relation between an identity and that person's sensitive information from the video. Therefore, the first step in privacy modeling is to determine the identity leakage. Let us start with the analysis of the human recognition system. We recognize people generally by their name, face, and habits etc. Table 4.1 enumerates usual idiosyncrasies used by human beings to recognize fellow humans. In our formalization of the identity leakage, we model 'face' as *who* evidence, 'gait' and 'social behavior' as *what* evidence, 'time' as *when* evidence, and spatial information as *where* evidence [SAM+10]. Other aspects such as 'associated objects' and 'who else they meet' can be considered by determining the number of objects present in the video.

We measure the identity leakage through the degree of anonymity. An identity leakage with κ-anonymity means that the size of the smallest group to which the adversary can associate the individual's identity is κ [Swe02]. With detection of each idiosyncrasy, we are able to associate the identity of the individual to a subgroup of people (the default initial group is the whole world population). For example, when it is detected that the place is 'Smart Lab, NUS', through prior knowledge we can relate the identity of the individuals in the video to the group of people who visit 'Smart Lab'. Further, if time is also detected as evening, we can use the knowledge that only half of them are expected to be in lab in the evening, reducing the association group

Table 4.1: Different idiosyncrasies which human being use in order to recognize other people.

Evidence	Idiosyncrasy	Example
who	Face	Face is considered to have features which distinguish people accurately.
what	Clothes	Depending on their taste, people repeat a particular style of clothes.
	Gait	Many people have a particular way of moving their body parts.
	Who else they meet	A person meeting a professor very frequently is probably one of his students.
	Social Behavior	This includes gender specific, culture specific, and religion specific behavior.
when	Timing of actions	What time they take lunch, what time they go office etc.
where	Spatial information	Person inside a particular shop is most likely the owner.

by half. Hence, the identity leakage depends on the prior knowledge of the adversary and corresponding observations from the video. We model the knowledge of an adversary as a rule based expert system [HRWL84]. An expert system contains facts and beliefs learned over time in its knowledge base which can be used to interpret the observation (in our case its purpose is to infer the identity from given evidences). The structure of knowledge base and its application will be discussed later. Note that a significant amount of work is being done on knowledge modeling in the interdisciplinary fields of Natural Language Processing, Information Retrieval, Machine Learning, and Knowledge Representation and Reasoning [VHLP08]. Systems are being developed which can learn these rules automatically [Fer10, FBRG11] and which can be used to cause the privacy loss.

Figure 4.1 shows the overall framework for the identity leakage calculation. Events are detected from the video of each camera and analyzed to enumerate the number of targets; and consequently an event list is constructed for each target. In the figure, $T_1, T_2, \ldots, T_m$ denote m targets and $E_1, E_2, \ldots, E_m$ denote the corresponding event lists. The event list contains all the events in which a particular target is detected. Overall identity leakage I_i for the target T_i is calculated by using anonymity based fusion of the information from the corresponding events in the event list E_i. The anonymity is calculated for each target appearing in the detected events using the adversary's knowledge base.

4.1.1 Video Segmentation

The video segmentation can be performed based on various criteria. For example, a segment can be the video between two consecutive background frames encompassing non background frames with motion and activity [LGW02]. Alternatively, video can be segmented based on the number of people [STT06]. The proposed framework is independent of how we segment the video. Event detection is applied on all the segments which contain people in it. A segment can result in multiple events.

4.1.2 Evidence Detection

After segmenting the video into events, we detect evidences from each event. This is done by analyzing the information present in the video. If the information is sufficient to recognize the place, we consider that the *where* evidence is detected. Similarly, if the event contains time information, it leads to *when* evidence. The *what* evidence is always present if the event involves one or more persons; which is true by definition.

4.1.3 Adversary Knowledge Base

An adversary can associate the events and activities to a person or group of persons only if s/he has appropriate prior knowledge. The prior knowledge of an adversary is represented using propositional logic statements. This knowledge can be obtained using machine learning techniques or it can be expert knowledge related to the application scenario. Every statement consists of a premise and conclusion. A premise is a proposition which is used as the foundation for drawing conclusions. In our case, each premise proposition statement consists of information about an event. An event consists of following attributes: *when*, *where*, and *what* (we exclude the *who* because face is assumed obscured in the video). Hence, a premise of a statement is represented by a 4-tuple $\mathcal{P}(t_s, t_e, act, loc)$; where t_s and t_e are the start and end time of event respectively (when), act is the activity (what), and loc is the location (where). For example, a premise statement $\mathcal{P}$(25-Oct-2010:10:35, 25-Oct-2010:11:10, “working”, “smart lab”) means “a group of people were working in the smart lab from 10:35 hrs to 11:10 hrs on 25-Oct-2010”. This premise leads to the corresponding conclusion $\mathcal{C}(grp)$, where grp is the associated group of people. Knowledge base consists of pairs of premise and conclusion statements. A knowledge

base entry $\mathcal{P} \Rightarrow \mathcal{C}$ denotes that if a premise $\mathcal{P}$ is true, it leads to conclusion $\mathcal{C}$. For example, $\mathcal{P}$(25-Oct-2010:10:35, 25-Oct-2010:11:10, "working", "smart lab")$\Rightarrow$ $\mathcal{C}$(G1) would mean that "The group of people who are working in the smart lab from 10:35 hrs to 11:10 hrs on 25-Oct-2010 are G1". Note that an absence of any attribute in the tuple is represented by a null symbol 'ϕ'. For instance, a premise statement $\mathcal{P}(\phi,\phi,$"dancing",$\phi)$ denotes "a group of people was involved in dancing activity in the scene". This premise leads to the corresponding conclusion $\mathcal{C}$(G2) "The group of people in the video who are involved in the dancing activity are G2".

Using the above propositional statements, we can build propositions for each type of idiosyncrasy listed in Table 4.1. For example, the identity leakage through clothes can be represented as $\mathcal{P}(t_s,\ t_e,$ "kurta", "office") to distinguish people who wear kurta in office, social behavior related the identity leakage can be represented as $\mathcal{P}(t_s,\ t_e,$ "praying", "temple") which can map to the people of particular religion who go to temple for praying, the proposition for gait related the identity leakage can be constructed as $\mathcal{P}(t_s,\ t_e,$ "hand in pocket", "temple"), and companion related the identity leakage as $\mathcal{P}(t_s,\ t_e,$ "with Mukesh", "temple").

The event contains the information that can be learned by the adversary about any individual. Every event is a potential source of the identity leakage. The identity leakage can take place in two ways:

- Through individual events in isolation i.e. considering each event as the only one event in the video. Every event has *what, when* and *where* information which can be used to associate a person's identity to a particular group of people.
- Through spatio-temporal sequence of events which might map to identity revealing patterns present in the knowledge base. The matching patterns further restrict the identity to a subgroup of people.

Although all statements of the knowledge base conclude in an association group, they differ for both the cases discussed above. The statements used for the identity leakage from events generally have propositions that are generated by only that event, whereas for later case the premise may consist of multiple propositions derived from multiple events which might be from multiple cameras. We will calculate both types of identity leakages (due to individual events and event patterns) and combine them to find the overall identity leakage.

4.1.4 Identity Leakage from Individual Events

The propositions generated in the previous section contain the adversary's knowledge base. However, the identity leakage only occurs if similar propositions are also found in the video. Therefore, for each proposition derived from the events, we find mapping in the knowledge base. If an appropriate mapping is found, we add the corresponding associated group in the set of mapped groups **G**. Let $\mathcal{G}^e$ be the resulting association group due to event information. It is calculated as the intersection of all the groups in **G** as following:

$$\mathcal{G}^e = \cap\{G \mid \forall G \in \mathbf{G}\} \tag{4.1}$$

For example, suppose the knowledge base consists of propositions $\mathcal{P}_1$ to $\mathcal{P}_{10}$ with the corresponding association groups G1 to G10; and the event under consideration generates propositions $\mathcal{P}_2$ and $\mathcal{P}_8$. In this case, the set of mapped groups **G** ={G2, G8} and the resulting association group $\mathcal{G}^e$ =G2∩G8. The above equation implies that using the information present in the event, we are able to associate the identity of the person seen in the video to a group of people $\mathcal{G}^e$.

4.1.5 Identity Leakage through Multiple Event Patterns

In the previous section we modeled the identity leakage by considering the events in isolation. However, the adversary may track the person over multiple events and multiple cameras which allows the adversary to exploit the knowledge of event patterns to further associate the person to a smaller group size. In order to model the identity leakage through event patterns, events are analyzed to detect the total number of targets present in the video. A separate event list is created for each target as shown in Figure 4.1. The target association across events is done based on the similarity of appearance like height, clothes etc. This is because the adversary can obtain event sequence only if the person looks similar across cameras.

Once we build the event list for all the targets, we fuse the identity leakage for each target using the information obtained from its associated event list. Let $\mathcal{G}^e_{ij}$ be the association group for the j^{th} event in the event list E_i, which corresponds to target T_i, and is calculated using Equation 4.1.

To understand how the knowledge of patterns helps in the identity leakage, let us consider

the following example. Suppose the adversary knows that G1= { A1, A2, ..., A10} people are expected at site 1, and G2= {A1, A2, B1, B2, ..., B6} people are expected at site 2. If A1 appears at site 1 and site 2 in two separate events (from separate cameras), the corresponding anonymities are 10 for event 1 and 8 for event 2. This results in an anonymity of 8 for A1 (i.e. minimum of the two values), if events are considered in isolation. On the other hand, even without facial information the adversary can identify that the person detected at site 1 as well as site 2 is the same through visual similarity. The adversary has the pattern information that only A1 and A2 are seen at both sites which results in reduced anonymity of 2. The pattern information (only A1 and A2 are seen at both sites) is embedded in G1 and G2 and it can be easily obtained by doing intersection of G1 and G2. Hence, the combined association group $\mathcal{G}'_i$ for target T_i is calculated as intersection of all the association groups of corresponding events in its event list E_i (See Equation 4.2).

$$\mathcal{G}'_i = \mathcal{G}^e_{i1} \cap \mathcal{G}^e_{i2} \cap \ldots \mathcal{G}^e_{in_e} \tag{4.2}$$

where n_e is the number of matching propositions for all the events in E_i.

In the discussion above, it is assumed that the adversary has the knowledge of all population in entirety. In practice the adversary may not have knowledge of G1 and G2 but may only know that if someone is seen at both site 1 and site 2, it is either A1 or A2, which can be stored in the knowledge base using following statement:

$$\mathcal{P}(\phi, \phi, \phi, Site1) \wedge \mathcal{P}(\phi, \phi, \phi, Site2) \Rightarrow \mathcal{C}(\text{A1, A2}) \tag{4.3}$$

Now, if there are two events generating propositions $\mathcal{P}$(ϕ,ϕ,ϕ,Site1) and $\mathcal{P}$(ϕ,ϕ,ϕ,Site2), they will not cause any identity leakage individually as they do not have any mapping statements. However, when found to be related to the same target, both the events can be mapped together to the pattern statement discussed above causing additional identity leakage. Let $\mathcal{G}^p_{i1}, \mathcal{G}^p_{i2}$... be the association groups for the matching pattern statements. The overall association group is now calculated by intersecting all the association groups corresponding to events ($\mathcal{G}^e_{ij}$) and event patterns ($\mathcal{G}^p_{ij}$):

$$\mathcal{G}_i = (\mathcal{G}^e_{i1} \cap \mathcal{G}^e_{i2} \cap \ldots \mathcal{G}^e_{in_e}) \cap (\mathcal{G}^p_{i1} \cap \mathcal{G}^p_{i2} \cap \ldots \mathcal{G}^p_{in_p}) \tag{4.4}$$

In the above equation, n_p is the number of matching patterns for all the events in E_i.

The anonymity, κ_i, of target T_i is calculated as size of overall association group $\mathcal{G}_i$:

$$\kappa_i = |\mathcal{G}_i| \tag{4.5}$$

Finally, the identity leakage, I_i, for target T_i is the inverse of the anonymity by definition and is computed as:

$$I_i = \frac{1}{\kappa_i} \tag{4.6}$$

If all the events belong to one person, that person should be common among all the association groups. Nevertheless, there can be multiple people satisfying the same set of propositions. For example, there can be multiple people who come to the lab at night and do similar activities. Therefore, the identity leakage is generally less than one.

4.2 Privacy Loss

The privacy loss takes place when the adversary acquires some sensitive information about the identified individual. In the previous model, a single sensitivity index is obtained for all individuals. In this chapter we enhance the model to calculate sensitivity index for each person in the video. Let W be the priority matrix defined as follows:

$$W = \{w_{ik} \mid i \in [1, m], k \in [1, l]; w_{i1} + w_{i2}, ... + w_{il} = 1\} \tag{4.7}$$

where the weights w_{ik} is set by the i^{th} individuals and reflects the individual's priority of the corresponding attribute a_k as sensitive information. For each target we detect the sensitive attributes in all of the events in the corresponding event list and build a sensitivity matrix as follows:

$$S = \{s_{ik} \mid i \in [1, m], k \in [1, l]\} \tag{4.8}$$

Each column is related to one target and rows to the sensitive attributes. The elements of

the array are calculated as follows:

$$s_{ik} = \begin{cases} 1 & \text{if } k^{th} \text{ attribute is detected for target } T_i; \\ 0 & \text{otherwise.} \end{cases}$$

The sensitivity index for each i^{th} target can be calculated as follows:

$$\Psi_i = \sum_{k=1}^{l} w_{ik} * s_{ik} \tag{4.9}$$

Now, if $I = \{I_1, I_2, ...I_m\}$ is the vector of the identity leakage probabilities, the privacy loss γ_i for the i^{th} individual can be calculated as:

$$\gamma_i = I_i \Psi_i \tag{4.10}$$

In order to reduce the privacy loss, we need to minimize both the identity leakage and the sensitive information. Interestingly, even if one of them is close to zero we can attain very low values of the privacy loss.

4.3 Experimental Results

To demonstrate the utility of the proposed model, we conduct three experiments. In the first experiment we highlight the difference between the identity leakage and the privacy loss. In the second experiment, the effect of multiple events on the identity leakage is shown. This is in contrast to our earlier work (Chapter 3) which determines the privacy loss based only on a single activity in the video. Finally in the third experiment, we show how the proposed framework is used to calculate the privacy loss in the case of a multi-camera surveillance video.

4.3.1 Experiment 1: Identity Leakage Vs Privacy Loss

The distinction between the identity leakage and the privacy loss is demonstrated using the following experiment. We consider a case where a person is sick and visits a hospital. He does not want his colleagues to know his disease and the doctor with whom he is consulting. Probably because the doctor's specialization might reveal the identity. Here, there are two

(a) (b)

(c) (d)

Figure 4.2: Four pictures take by surveillance cameras placed around an hospital.

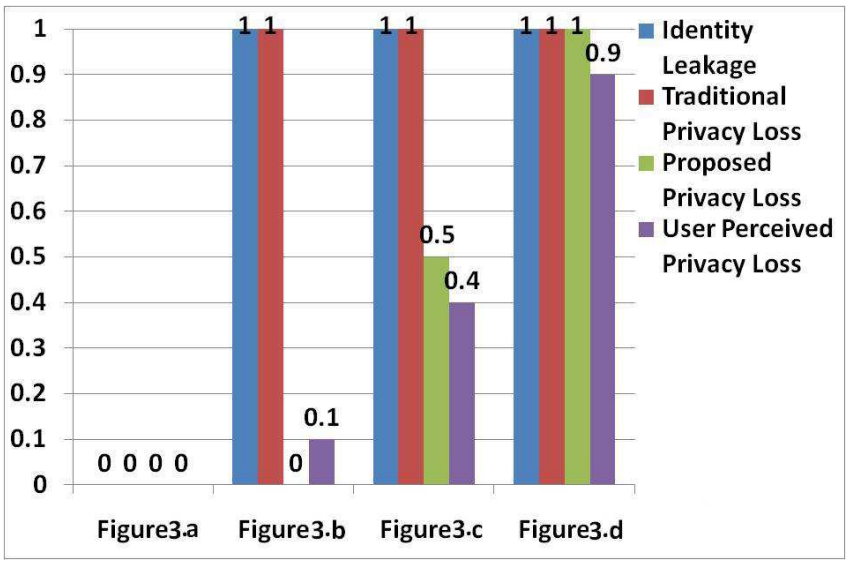


Figure 4.3: Identity Vs Privacy.

sensitive attributes $A = \{a_1, a_2\}$ where a_1 is companion and a_2 is location. The corresponding priority weights for the sensitive information are given as $W = \{w_1, w_2\}$ where $w_1 = 0.5$, and $w_2 = 0.5$. He goes to the hospital and the surveillance camera records four separate images of the person. For this example we assume that each image is representative of one event. The four images are shown in Figure 4.2. The sensitivity matrix is a row vector since we are analyzing the privacy loss of only the patient, which is denoted as $S = \{s_1, s_2\}$. The values of s_1 and s_2 are different for different images; and they are determined below.

In Figure 4.2(a), we cannot see the person's face, hence the privacy loss predicted by traditional models is 0. The picture also does not have any other information which can be used for implicit inference channel, hence proposed model also gives zero identity leakage as well as privacy loss. In Figure 4.2(b) we can see the person's face implying a privacy loss of 1 with traditional models of privacy. However, since no sensitive information is available in this image, s_1 and s_2 are taken as 0, which results in zero privacy loss using the proposed method (Equation 4.10). Figure 4.2(c) clearly has companion information, making $s_2 = 1$. This results in privacy loss of $\gamma = 0.5 * 1 = 0.5$. Finally, from 4.2(d) we can exactly find out the disease through hospital name. Hence privacy loss $\gamma = 1 * (0.5 * 1 + 0.5 * 1) = 1$. Figure 4.3 shows the results of the experiment. To get the user perceived privacy loss, five students aged between 20 to 30 were explained the situation and they were asked to rate the privacy loss for each image between 0 to 1.

Table 4.2: Knowledge base for experiment 2.

1. $\mathcal{P}(\phi,\phi,\text{SL})\Rightarrow$G1(A1-10)	3. $\mathcal{P}(\text{EV},\phi,\text{SL})\Rightarrow$G3(A3-4,A7,A9)
2. $\mathcal{P}(\phi,\text{DC},\text{SL})\Rightarrow$G2(A1-2,A4-7)	4. $\mathcal{P}(\phi,\text{RN},\text{SL})\Rightarrow$ G4(A5-7,A9-10)

Table 4.3: Event description of the video for experiment 2.

Event	Description	Event	Description	Event	Description
C_1	T_1, WK	C_4	T_1, WK	C_7	T_2, RN
C_2	T_1, WK	C_5	T_1,T_2, DC	-	-
C_3	T_2, WK	C_6	T_3, WK	-	-

There are two implications of the results. Firstly, if the video does not contain any sensitive information for the person, we do not need to hide the identity information. The video will not cause any privacy loss. On the other hand, if identity cannot be inferred from the video, the video can be released with all the sensitive information. Although this is generally not possible as sensitive information itself can cause the identity leakage through other implicit inference channels.

4.3.2 Experiment 2: Event Based Identity Leakage

The goal of this experiment is to highlight the effect of multiple events on the identity leakage. We use a video clip from a single camera recorded in a lab scenario. The knowledge base. In this case has statements shown in Table 4.2. The video, half an hour in length, consists of seven events described in Table 4.3 with three targets involved. Here, SL means 'Smart Lab' and for clarity we mention the events as $C_1, C_2, \ldots$, etc., which later form the event lists E_i. The representative images from four of them are shown in Figure 4.5. In this video clip, all three types of evidence are detected as follows: *what*= walking (WK), running (RN), discussion (DC), *where*= smart lab (SL), *when*= evening (EV). Since all the events had starting and ending time same as evening, in propositions we mention both start and end time using single EV.

Since the proposition generation depends on the targets, we describe it as event list for individual targets. In Table 4.4 the first column shows the events in which the target is detected. The second column shows the proposition generated by that event. In the third column we write the association group according to the mapping proposition in the knowledge base shown in Table 4.2. We calculate the final association group by calculating the intersection of all groups due to individual events. Table 4.4 shows the results for T_1 and T_2. Target T_3 only appears in one event and generate the proposition $\mathcal{P}$(EV, WK, SL) which map to statement 3 to give $\mathcal{G}_3 =$

Table 4.4: Event lists and identity leakage for individual targets.

E_1 (Target T_1)			E_2 (Target T_2)		
C_1	$\mathcal{P}$(EV,WK,SL)	G3	C_3	$\mathcal{P}$(EV,WK,SL)	G3
C_2	$\mathcal{P}$(EV,WK,SL)	G3	C_7	$\mathcal{P}$(EV,RN,SL)	G4
C_4	$\mathcal{P}$(EV,DC,SL)	G2	C_4	$\mathcal{P}$(EV,DC,SL)	G2
	$\mathcal{P}$(EV,WK,SL)	G3		$\mathcal{P}$(EV,WK,SL)	G3
C_5	$\mathcal{P}$(EV,WK,SL)	G2	-	-	-
$\mathcal{G}_1 = G2 \cap G3 = (A7, A9)$			$\mathcal{G}_2 = G2 \cap G3 \cap G4 = (A7)$		

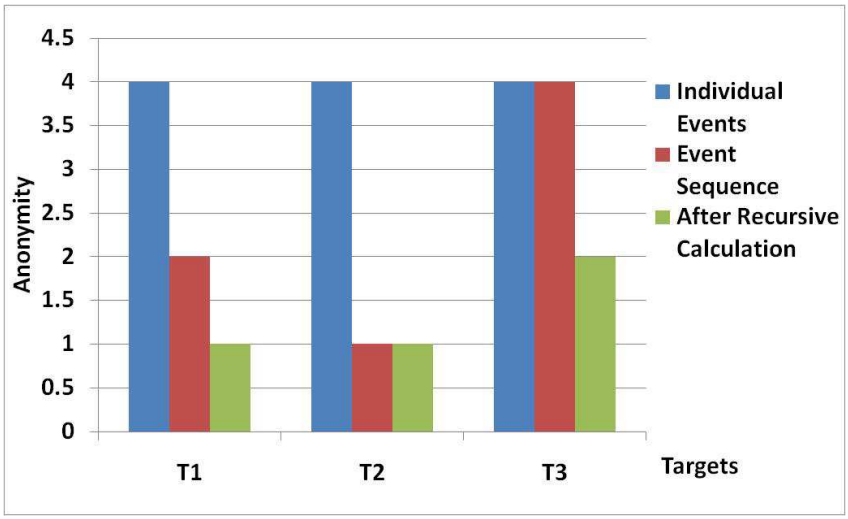


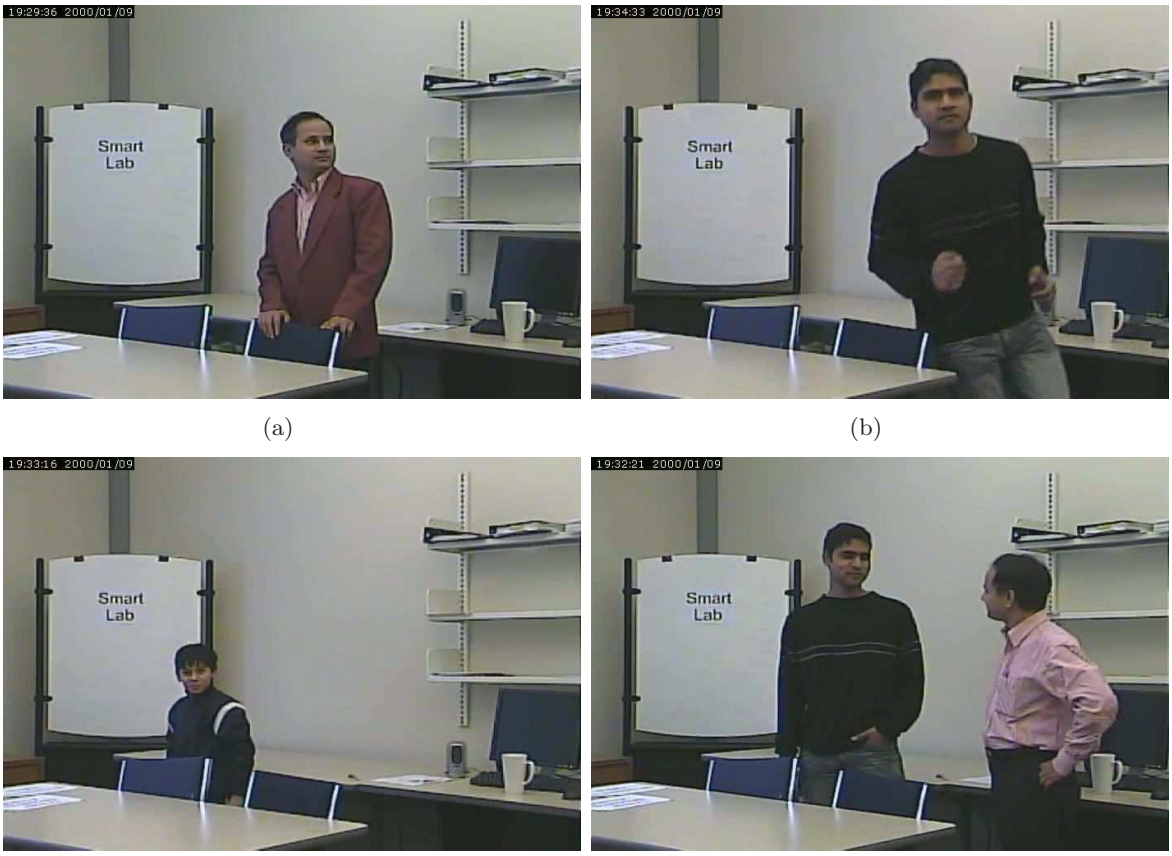
Figure 4.4: The anonymity when we consider events in isolation and event sequences. The third bar shows the results of recursive identity leakage.

G3. In case the event generated proposition does not have mapping in the knowledge base, the association group is assumed to be universal group (UV) which is superset of all other groups. Note that if only individual events are considered, the association groups would be the smallest from the list, which is incidently the same for all three targets i.e. 4. Figure 4.4 compares anonymities calculated using both methods.

It is interesting to observe that the identity leakage can produce regenerative effect which can result in increased identity leakage. For example, in this experiment we conclude that the anonymity of T_1 is anonymous between A4 and A7 and T_2 is known to be A7. Since we know that A7 is T_2, we can conclude that T_1 is A4. The exact identities of T_1 and T_2 also reduce the anonymity of T_3 to two (A3 or A9). The ground truth is A4, A7, A9.

4.3.3 Experiment 3: Privacy Loss from Multiple Cameras

If the adversary has access to multiple camera videos where the same person is spotted at multiple places, the adversary can use the knowledge of spatiotemporal patterns to infer the



(a) (b)

(c) (d)

Figure 4.5: Representative images from four events of the video recoded in smart lab.

(a) DE (b) AL

(c) SC (d) IC

Figure 4.6: Representative images from four cameras: (a) Department Entrance, (b) Audio Lab, (c) Staff Club, (d) Canteen.

Table 4.5: Knowledge base for experiment 3.

Statements for individual events.	
1. $\mathcal{P}(\phi,\phi,SC)\Rightarrow$G1(B1-10,C1-2,A1-3)	5. $\mathcal{P}(EV,\phi,SC)\Rightarrow$G5(A1-2,B1-5,C1-2)
2. $\mathcal{P}(\phi,\phi,AL)\Rightarrow$G2(A1-10,B1-5,D1-2)	6. $\mathcal{P}(EV,\phi,AL)\Rightarrow$G6(A1-4,B1-3,D1-2)
3. $\mathcal{P}(\phi,\phi,IC)\Rightarrow$G3(A1-5,B1-5,C8-10)	7. $\mathcal{P}(EV,\phi,IC)\Rightarrow$G7(A1-5,B1-3,C8)
4. $\mathcal{P}(\phi,DC,\phi)\Rightarrow$G4(A1-3,B1-4,D5-8)	-
Statements for multi-event patterns.	
8. $\mathcal{P}(\phi,\phi,DE)\wedge\mathcal{P}(\phi,\phi,IC)\Rightarrow$G8(A1,A8,B1,B8)	
9. $\mathcal{P}(\phi,\phi,DE)\wedge\mathcal{P}(\phi,\phi,SC)\Rightarrow$G9(A1,B1-3, C1-3)	
10. $\mathcal{P}(\phi,\phi,DE)\wedge\mathcal{P}(\phi,\phi,AL)\Rightarrow$G10(A1-6,B1-6, D1)	

Table 4.6: Description of events captured by cameras.

Camera 1 - Department Entrance (DE)					
Event	Description	Event	Description	Event	Description
C_{11}	T_1, WK	C_{16}	T_4, WK	C_{112}	T_4, WK
C_{12}	T_1, WK	C_{17}	T_6, WK	C_{113}	T_2, RN
C_{13}	T_5, WK	C_{18}	T_5, WK	C_{114}	T_2, WK
C_{14}	T_2, WK	C_{19}	T_2, WK	-	-
C_{15}	T_3, WK	C_{110}	T_3, WK	-	-
Camera 2 - Audio Lab (AL)					
Event	Description	Event	Description	Event	Description
C_{21}	T_1,T_2, DC	C_{23}	T_1, WK	C_{25}	T_1, RN
C_{22}	T_6, RN	C_{24}	T_2, WK	-	-
Camera 3 - Staff Club (SC)					
Event	Description	Event	Description	Event	Description
C_{31}	T_4, WK	C_{35}	T_5, WK	C_{39}	T_5, WK
C_{32}	T_3, WK	C_{36}	T_5, WK	C_{310}	T_3, WK
C_{33}	T_4, WK	C_{37}	T_4, WK		
C_{34}	T_3, T_1, DC	C_{38}	T_5, WK		
Camera 4 - Canteen (IC)					
Event	Description	Event	Description	Event	Description
C_{41}	T_1, WK	C_{43}	T_4, WK	C_{45}	T_4, WK
C_{42}	T_1, WK	C_{44}	T_1, WK		

identity. In this experiment we demonstrate two main contributions: (1) The effect of multiple cameras on the identity leakage and (2) The privacy loss assessment in a multi-camera scenario. For this experiment we recorded video at four places in the department building: (1) Department Entrance (DE) (2) Audio Lab (AL) (3) Staff Club (SC) and (4) Canteen (IC) (Figure 4.6). A total of 40 people are expected in the department (A1-10, B1-10, C1-10, D1-10). However, the adversary does not have knowledge about all of them. The adversaries knowledge is limited to the propositional logic statements given in Table 4.5.

Six actors created a series of events at these sites. Table 4.6 provides the description of all the events captured at these sites. These actors were involved in one of the following activities: discussion (DC), walking (WK), and running (RN). For calculation of anonymity, we created event lists for each target separately. The time was detected as evening (EV) in all the cameras except camera 1 at the department entrance.

The event lists for the targets and generated propositions are given in Tables 4.7. Similar

Table 4.7: Event lists and identity leakage for targets.

E_1 (Target T_1)					
C_{11}	$\mathcal{P}(\phi$,WK,DE)	UV	C_{41}	$\mathcal{P}$(EV,WK,IC)	G7
C_{12}	$\mathcal{P}(\phi$,WK,DE)	UV	C_{42}	$\mathcal{P}$(EV,WK,IC)	G7
C_{112}	$\mathcal{P}(\phi$,WK,DE)	UV	C_{44}	$\mathcal{P}$(EV,WK,IC)	G7
C_{21}	$\mathcal{P}$(EV,DC,AL)	G4,G6	Pattern	8	G8
C_{23}	$\mathcal{P}$(EV,WK,AL)	G6	Pattern	9	G9
C_{25}	$\mathcal{P}$(EV,RN,AL)	G6	Pattern	10	G10
C_{34}	$\mathcal{P}$(EV,DC,SC)	G4, G5			
$\mathcal{G}_1$ = G4∩G5...G10= (A1,B1)					
E_2 (Target T_2)					
C_{14}	$\mathcal{P}(\phi$,WK,DE)	UV	C_{21}	$\mathcal{P}$(EV,DC,AL)	G4,G6
C_{19}	$\mathcal{P}(\phi$,WK,DE)	UV	C_{24}	$\mathcal{P}$(EV,WK,AL)	G6
C_{113}	$\mathcal{P}(\phi$,RN,DE)	UV	Pattern	10	G10
C_{114}	$\mathcal{P}(\phi$,WK,DE)	UV			
$\mathcal{G}_2$ = G4∩G6∩G10= (A1-3,B1-3)					
E_3 (Target T_3)					
C_{15}	$\mathcal{P}(\phi$,WK,DE)	UV	C_{34}	$\mathcal{P}$(EV,DC,SC)	G4, G5
C_{110}	$\mathcal{P}(\phi$,WK,DE)	UV	C_{310}	$\mathcal{P}$(EV,WK,SC)	G5
C_{32}	$\mathcal{P}$(EV,WK,SC)	G5	Pattern	9	G9
$\mathcal{G}_3$ = G4∩G5∩G9= (A1,B1-3)					
E_4 (Target T_4)					
C_{16}	$\mathcal{P}(\phi$, WK,DE)	UV	C_{43}	$\mathcal{P}$(EV,WK,IC)	G7
C_{31}	$\mathcal{P}$(EV,WK,SC)	G5	C_{45}	$\mathcal{P}$(EV,WK,IC)	G7
C_{33}	$\mathcal{P}$(EV,WK,SC)	G5	Pattern	8	G8
C_{37}	$\mathcal{P}$(EV,WK,SC)	G5	Pattern	9	G9
$\mathcal{G}_4$ = G5∩G7∩G8∩G9= (A1,B1)					
E_5 (Target T_5)					
C_{13}	$\mathcal{P}(\phi$,WK,DE)	UV	C_{38}	$\mathcal{P}$(EV,WK,SC)	G5
C_{18}	$\mathcal{P}$(EV,WK,DE)	UV	C_{39}	$\mathcal{P}$(EV,WK,SC)	G5
C_{35}	$\mathcal{P}$(EV,WK,SC)	G5	Pattern	9	G9
$\mathcal{G}_5$ = G5∩G9= (A1,B1-3, C1-2)					
E_1 (Target T_6)					
C_{17}	$\mathcal{P}(\phi$,WK,DE)	UV	Pattern	10	G10
C_{22}	$\mathcal{P}$(EV,RN,AL)	G6	-	-	-
$\mathcal{G}_6$ = G6∩G10= (A1-4,B1-3,D1)					

to previous experiment, the first column tells the events in which the target was detected. The second column tells the proposition generated by the event. In the third column we write the association group derived from the knowledge base by proposition mapping. An event proposition may have multiple matches in knowledge base, in that case we list all the groups in the third column.

In order to calculate the privacy loss, we need to determine the sensitive information matrix. In this experiment we have chosen the sensitive attributes to be: (1) Companion (2) Running activity (3) Height (4) Clothes. For the given set of participants, we got similar priorities for the sensitive attributes as given in the priority matrix W below:

$$\begin{pmatrix} 0.45 & 0.30 & 0.15 & 0.10 \\ 0.45 & 0.30 & 0.15 & 0.10 \\ 0.45 & 0.30 & 0.15 & 0.10 \\ 0.45 & 0.30 & 0.15 & 0.10 \end{pmatrix} \tag{4.11}$$

The weights have been determined based on the common notions of privacy. People are more sensitive to their company than their clothes or height. Similarly, they might not feel comfortable being watched while running. The sensitivity matrix S can be derived from event descriptions as follows:

$$\begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix} \tag{4.12}$$

Since $I_i = 1/|\mathcal{G}_i|$, the identity leakage vector is calculated as:

$$I = \{0.50, 0.17, 0.25, 0.5, 0.17, 0.14\}$$

With these values of W, S, and I, we can calculate the privacy loss. Figure 4.7 shows the resulting identity leakage and privacy loss in three scenarios: (1) Individual events (2) Patterns among single camera events (3) Patterns among multiple camera events. It can be seen that when multiple camera video is available and adversary has the knowledge of patterns, the identity leakage and the privacy loss increases.

Similarly, Figure 4.8 shows the identity leakage and the privacy loss for all the targets measured using proposed framework for multi-camera video. The identity leakage for T_1 is the highest because T_1 was seen at all four sites and was involved in all the activities, walking, running and discussion. T_2 and T_4 appear in the same number of events, however, T_4 appears in events from multiple cameras hence its identity leakage is higher from T_2. T_1 and T_4 have the same identity leakage, still the privacy loss of T_1 is higher than T_4. This shows the effect of the sensitive attributes on the privacy loss. For T_1, all sensitive attributes are detected whereas for T_4, only two out of four attributes are detected.

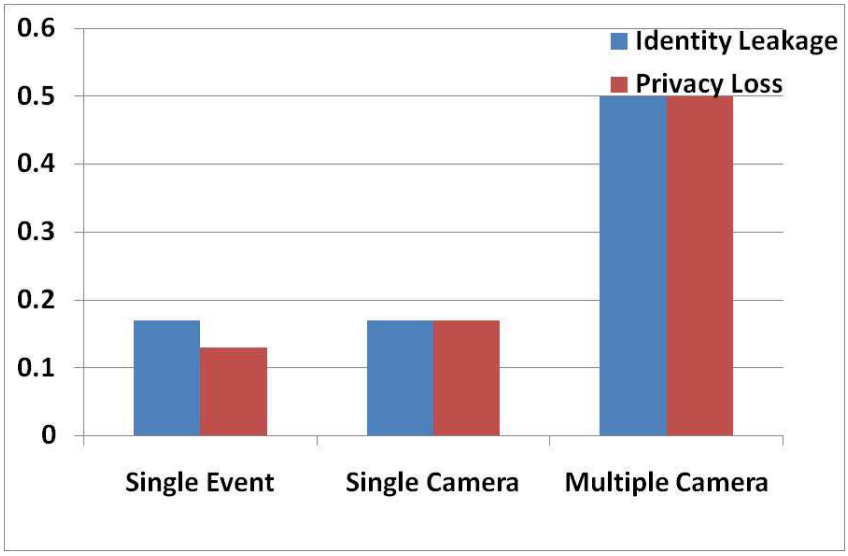


Figure 4.7: Identity leakage and privacy loss for T_1.

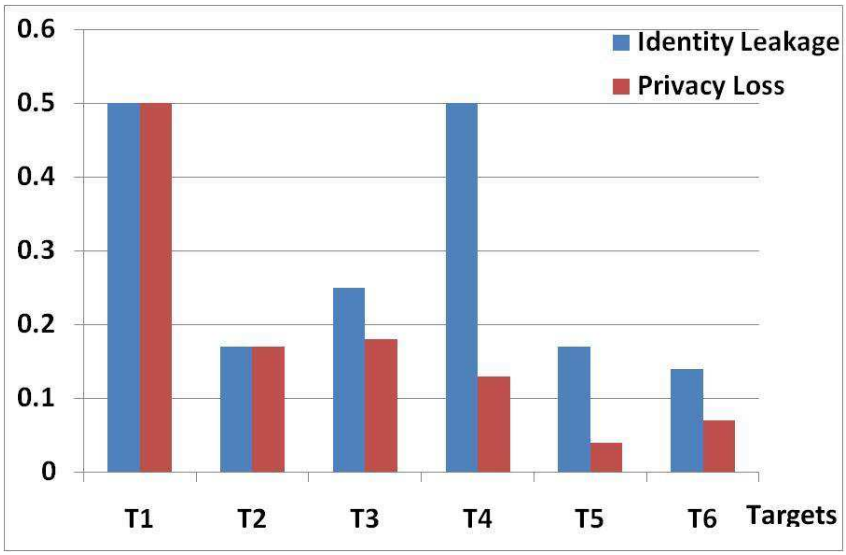


Figure 4.8: Identity leakage and privacy loss for all targets.

4.4 Discussion

This chapter highlights the privacy breach that exists in current privacy preserving methods which are based on only facial and appearance based cues, particularly for multi-camera video. Although the current technology is not robust enough to decipher the event information accurately, a human observer can definitely detect *what, when,* and *where* information from multi-camera video which may lead to privacy loss. In experiments it is demonstrated that even when the bodily cues of the identity are absent, in extreme cases the adversary can identify the individuals with small value of anonymity.

We acknowledge that the success of any privacy preserving method depends on the automated detection techniques and there is a whole body of researchers working on improving these detectors [DvGN+08], [PF11], [FBRG11]. In this chapter we restricted our focus on analyzing the various channels (other than the human face) that can cause identity leakage which we believe is an important and the first step towards future privacy-aware systems.

4.5 Conclusion

Privacy analysis is very important for any legitimate use of video data. In this work we provide a framework for calculating the privacy loss that might occur given public access of the multi-camera surveillance video. We recognize that the privacy loss occurs when adversary is able to map the identity of individuals to the sensitive information present in the video. The experimental results show that the privacy loss generally increases with the number of events in a video as well as the number of cameras. The proposed method can be configured for any adversarial knowledge and the sensitive attributes of the people making it flexible and applicable to a variety of applications. For example, in a surveillance scenario, the habitants of the surveilled premises can provide the sensitive attributes and the person, who has access to this surveillance video, can be considered as an adversary.

Chapter 5

Anonymous Surveillance

So far the main focus of thesis has been analyzing various channels through which a video can cause identity leakage and privacy loss. In this chapter we apply the research findings to traditional surveillance systems and analyze the privacy loss. Based on the privacy analysis, we propose a novel privacy protection framework that can provide robust measures of the privacy with minimal compromise in the surveillance utility.

We find that it is almost impossible to hide the identities in traditional surveillance systems. The local CCTV operators generally not only have a good knowledge of the surveillance site and current time, they are very familiar with the people who regularly appear in the video and hence have a good chance to identify them. Further, our study reveals that it is very hard to hide the *when* and *where* information from a local operator; therefore the implicit channels cause significant privacy loss.

To counter such contextual knowledge of the local CCTV operator, we propose an anonymous surveillance framework. The proposed framework advocates for context decoupling and recommends surveillance video feeds to be watched at a remote place. Before transmitting the video to a remote site, we anonymize the video such that the location where the video was recorded and the time of recording cannot be learned. At the remote place, we perform random re-assignment of the cameras to minimize the probability of remote operator building context by continuously watching the same video feed. User study also reveals that people are more comfortable when the video is watched remotely and the cameras are shuffled periodically.

In order to build an effective anonymous surveillance system, we need to address a series

of challenges into multidisciplinary areas of video streaming, networking, computer vision, and signal processing. Along with a detailed description of the novel privacy preserving framework, in this chapter we are also able to address two important challenges: bandwidth friendly background anonymization and workload equalizing camera assignment.

The main contributions of the chapter are following:

- We analyzed the privacy breach in video surveillance systems that exists due to prior knowledge of the CCTV operator.
- We proposed an anonymous surveillance framework that decouples the prior knowledge from the video and provides more robust privacy.
- We quantify the privacy loss that occurs through background video information and propose network friendly background anonymization method.
- We propose a workload based random assignment scheme of the cameras to remote operators that equalizes the workload distribution.

5.1 Chapter Organization

The rest of the chapter is organized as follows: a detailed analysis of the privacy loss in current surveillance systems is provided in Section 5.2. Different conclusions are derived from the discussion as driving forces for the proposed framework. Next we propose the anonymous surveillance framework in Section 5.3 and list out various challenges involved. During the description, we provide pointers to the current works in other areas that can be used for implementation of the system and find two important problems that still need to be addressed. The problem of background anonymization is dealt with in Section 5.4 and a workload based camera assignment is proposed in Section 5.5. We summarize the Chapter in Section 5.6.

5.2 Privacy Analysis

Video surveillance has been proven to be an effective means of ensuring security and safety. In video surveillance, cameras are placed at important places and video feeds are sent to local security offices. Because the automated surveillance is still in its infant stage, most of the

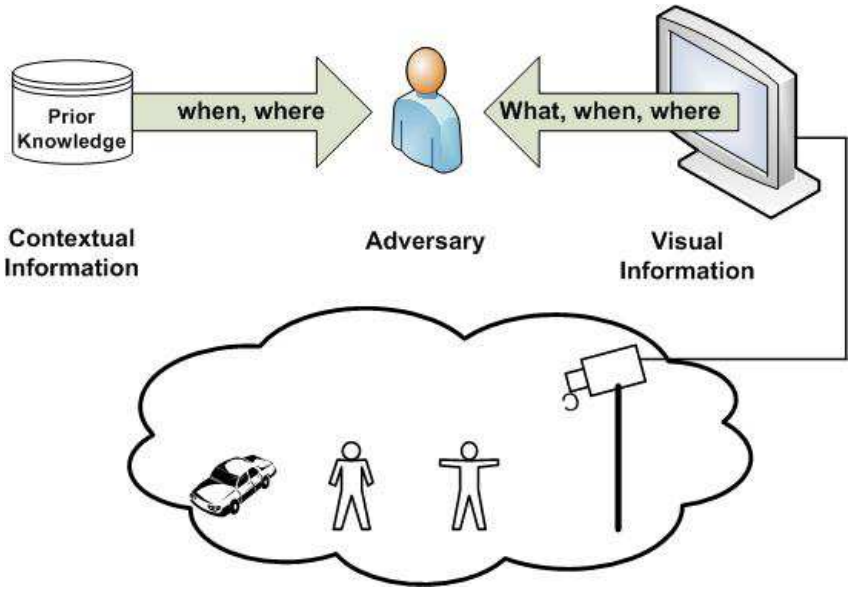


Figure 5.1: The *when* and *where* information can come from both video data and adversary's prior knowledge.

time these feeds are manually monitored by CCTV operators. This can cause privacy loss if the CCTV operator acts as an adversary and attempts to identify the individuals in the video, and acquire additional knowledge about them e.g. habits, company, and other lawful activities. Hence, it is important to protect the association of identity of individuals to the sensitive information in the video to reduce the privacy loss. In this section we first discuss various observations made based on the previously proposed privacy models, and then describe a user study that advocates anonymous surveillance for robust privacy.

5.2.1 Observations

Following are our observations regarding the privacy loss in traditional surveillance systems.

Observation # 1: Sensitive information cannot be removed

It is noticed in the privacy modeling (Chapter 3 and 4) that the privacy loss is association of the identity with the sensitive information. The very first approach of providing the privacy protection could be to remove sensitive information from the video. There are two problems with this approach which make it infeasible: (1) it is generally hard to automatically detect the sensitive information (2) the sensitive information is generally important for surveillance purposes, therefore removing all the sensitive information may render the data useless. The

second problem arises because both sensitive incidents and suspicious incidents have the common characteristics of high entropy; and these are difficult to distinguish. An incident removed from the video as sensitive incident has high probability of containing suspicious information. For example, a person may be sensitive to the objects s/he carries, but these objects may be weapons. Therefore, the identity leakage is a more feasible solution of the privacy preservation in surveillance systems.

Observation # 2: *who* information can be hidden through computer vision

We have discussed in previous chapters that the identity leakage occurs through events which mainly consist of four aspect: *who, what, when*, and *where*, which are called evidences. While *who* evidence uniquely identifies a person e.g. facial information, *what, when* and *where* information can be used to associate a person to a specific group of people depending on the obtained evidences and adversaries knowledge. In this chapter we assume that *who* evidence can be removed using computer vision techniques, or in the worst case the whole image can be globally transformed to hide the facial information. Even if the face hiding technique is not accurate, the presence of faces in the video does not cause privacy loss if the adversary is not able to identify them. The probability of face identification can be minimized if the viewer is not familiar with the faces of people under surveillance.

Observation # 3: *what* information is more important for surveillance and does not cause much privacy loss when detected alone in isolation

Hiding the identity leaking *what*, *when*, and *where* information from the local CCTV operator is very hard, if not impossible. Further, it is not adequate to remove *what* information from the video, because this will defeat the basic purpose of surveillance. One needs to detect the suspicious events and activities in order to assess the security threats, which requires the *what* information to be present in the video. However, the majority of surveillance tasks do not require a person's identity to be known e.g. intrusion, fight, quarrel, and theft. A list of important security threats and and corresponding surveillance tasks is provided in Table 5.1. Unlike *what* information, *where* and *when* information is less crucial for surveillance; and it can be easily anonymized by providing rules like "At this place people are not allowed to do activity x" or

Table 5.1: The surveillance task and associated security threat.

Task	**Respective security threat**
Change Detection	Vandalism, Camera tampering, Graffiti
Object Introduced	Abandoned baggage, Illegal Parking
Object Removed	Theft, Museum surveillance
Direction of Motion	Counter flow, One way, Immigration
Movement from A to B	Intrusion, Illegal turns, Walking patterns
Cross a zone multiple times	Car surfing, Loitering, Counting
Loitering in a zone	Loitering in a crowd
Overcrowding	Stampede at platforms and Ticket halls
Congestion	Traffic jam
Sound Detection	Aggression Fight, Assault
Sound Detection	Gun Shot Firearms
Sound Detection	Panic Scream, Shouts, Cry for help
Sound Detection	Vehicle Motorcycle, Lorry, Tank, Airplane
Counting	Vehicles and People Venue occupancy, Flow rates, Queue management
Boundary	Perimeter breaches, fence surveillance
Target count	Tailgating, Queue length
Quick movements	Quarrel, Fight
Object association	Gun, Knife, or other Weapons
Quick crowd movements	Stampede, Emergency evacuation
Smoke	Fire, Illegal smoking

"In this camera view people are not supposed to stand for more than 5 seconds". Fortunately, *what* alone does not cause any significant privacy loss [SAM+10] if *when* and *where* information is not present. Therefore we turn our focus to block *when* and *where* information.

Observation # 4: *what* and *when* information is generally available to the CCTV operator as prior knowledge

As mentioned earlier, video feeds are generally watched locally and the CCTV operator is provided with the location information of the video. Getting the current time is a trivial task as the operator and the camera are in the same time zone. The operator can associate this *when* and *where* information with the *what* information in the video and infer the identities of the people in the video, even when bodily cues are hidden. In order to protect this identity leakage, the operators should not be provided with *when* and *where* information which is very hard in traditional surveillance systems.

Observation # 5: It is very hard to hide *when* and *where* information from the video due to familiarity of the CCTV operator with the surveillance site

Even when the CCTV operator is not explicitly provided with the location information, s/he can easily infer it from video. It is very hard to obfuscate the video to hide the location information

because the CCTV operators are generally very familiar with the surveillance locality. Hiding text legends, symbols, or logos is easier but it is not effective in the case of traditional surveillance systems. To effectively hide the location information from the local operator, a large portion of the image needs to be hidden which is not suitable for surveillance application. In the user study described next, we found that if the video is monitored remotely by a CCTV operator who have minimal contextual overlap with the surveillance site, the above mentioned methods are sufficient for hiding location information.

5.2.2 User Study #1

In this study, the users were asked to play the role of an adversary and they are provided with three video clips. For each video clip, the users were asked to answer a set of questions that were designed to test the user's ability to learn *who, when* and *where* information.

The users were divided into two groups: the first group consisting of ten people and the second group had seven users. While the first group belonged to National University of Singapore (NUS), the second group had users from all over the world who had never been to NUS or Singapore. The video clips were recorded in NUS; therefore, in the rest of the paper we will refer to the first group of students as local group (ten people) and the second group of users as remote group (seven people). Both groups were allowed to watch the three video clips and answer questions. The users were kept unaware of the location where the video is recorded and the time of recording.

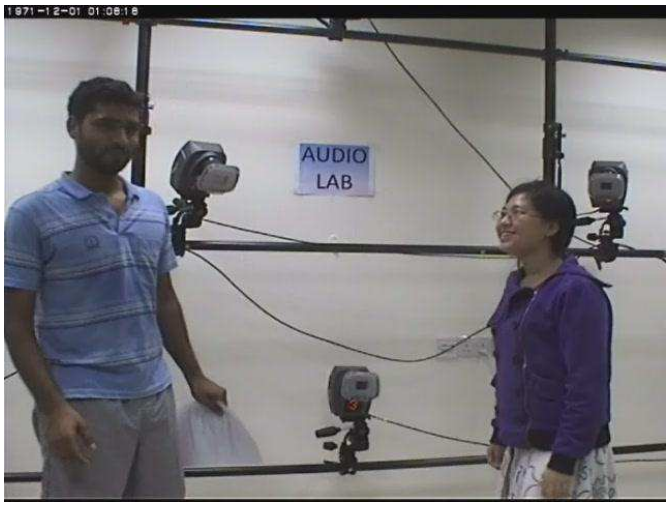
Video data set

Three video clips were used for the user study. All three videos were recorded in our university, though, in three different settings. The first video was recorded at the entrance of our department, and featured six people in it. In the second video, we captured an outdoor scenario. The video was recorded near the emergency exit of the same department with four actors performing various activities. The third video was recorded inside a lab with four people performing various activities. Figure 5.2 shows the representative frames from the three video clips.

Before presenting to the user, the video clips are transformed to remove the *when* and *where* information. The image regions which were revealing the place and time information were

(a)

(b)



(c)

Figure 5.2: The representative frames from the three video clips. Images from video clip 1 and 2 (i.e. a and b) have been modified to hide the university information.

(a)

(b)

(c)

Figure 5.3: The transformed representative frames from the three video clips.

Table 5.2: Questionnaire for user study #1

No.	Question
Q1	Do you recognize any person in the video?
Q2	Do you recognize any person by name?
Q3	Do you recognize the continent where the video is recorded?
Q4	Do you recognize the country where the video is recorded?
Q5	Do you recognize the city where the video is recorded?
Q6	Do you recognize the premises where the video is recorded?
Q7	Can you differentiate the time in the video among morning, noon, afternoon, evening, night?
Q8	Are you able to identify the time with the precision of hours?
Q9	According to the given definition of privacy loss, do you think any privacy loss has occurred in this video? Rate between 0 and 10 (0: no privacy loss, 10: full privacy loss).
Q10	According to the given definition of utility loss, do you think any utility loss has occurred in this video? Rate between 0 and 10 (0: no utility loss, 10: full utility loss).

pixelized to hide these evidences. The frames of the transformed video are shown in Figure 5.3.

Questions

The questionnaire consisted of ten questions in total, as shown in Table 5.2. The questions were divided into four groups: two *who* evidence related questions (Q1 and Q2), four *where* evidence related questions (Q3 to Q6), two *when* evidence related questions (Q7 and Q8), and two questions to record their overall feeling of the privacy loss[1] and the utility loss[2] (Q9 and Q10).

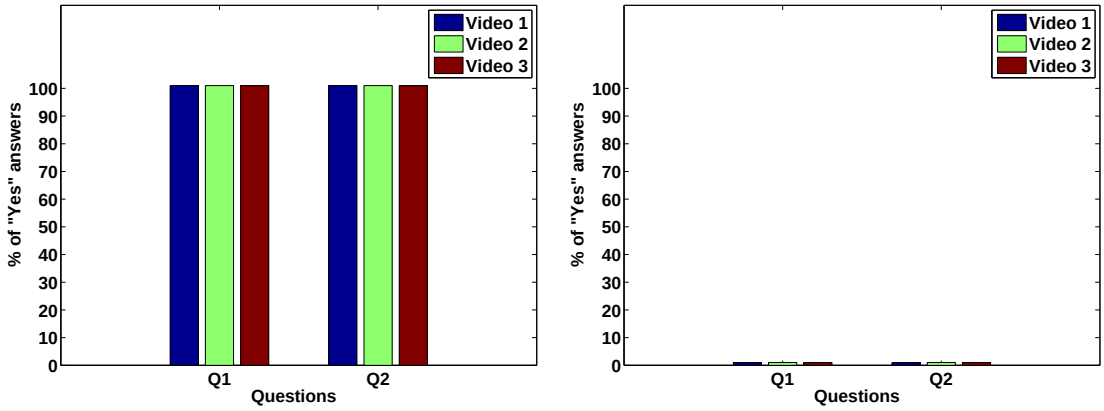
The answers to first eight questions were to be given in YES or NO. If the user gave YES answer, they were asked to specify. For example, if a user answered YES to Q3, we requested him to provide the name of the continent. The answer is counted as YES only if the specified answer is indeed correct, otherwise it is considered NO. The questions Q9 and Q10 were to be answered after they have finished watching the whole video and have answered all the questions. In these questions, the users were asked to give a rating for the privacy loss and the utility loss in the scale of 0 to 10.

Procedure

We asked the participants to play the role of an investigator with the goal of identifying people, place, and time in the video. The participants are allowed to pause and play the video as many times as needed. The questions in Table 5.2 were designed mainly to test the participant's

[1]Privacy loss is the measure of the who, where, when and what aspects of the information that can be gained from the video data [SAM+10]

[2]Utility loss of the video data refers to the decrease in the degree of accuracy by which security tasks can be accomplished.



(a) Local Participants (b) Remote Participants

Figure 5.4: User study results for questions 1 and 2.

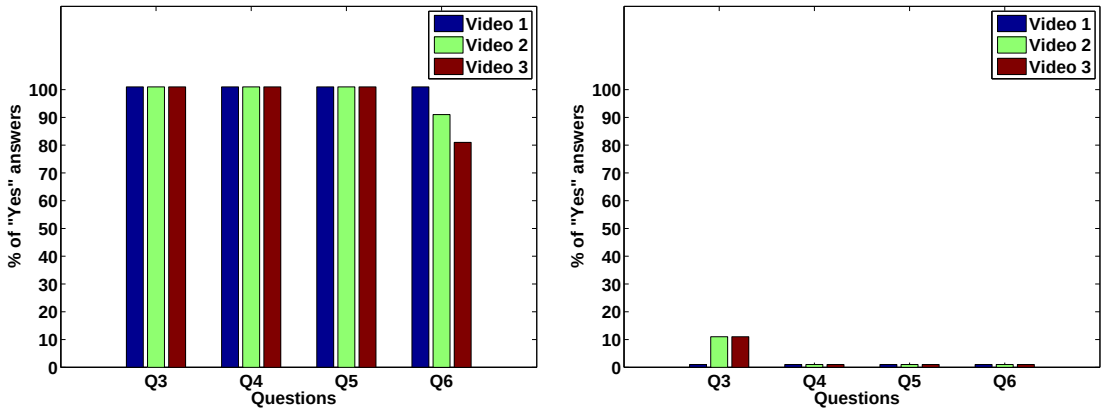
ability to judge who the people are in the video, what the place of recording is, and what the time of recording is.

Analysis of *who* Evidence

The faces were visible in all three videos. As shown in Figure 5.4, when local users watched the video clips, they could easily recall most of the people in the video. This is generally the case in current CCTV surveillance systems. However, when the same videos are shown to remote users, as expected, they could not recognize the individuals in the video because they did not have enough context or prior knowledge to recognize them. Hence, the prior knowledge decoupling performed in the proposed framework could preserve the identities even when the faces were visible. Note that this is important when the face detectors are not fully reliable and complete removal of facial information is hard.

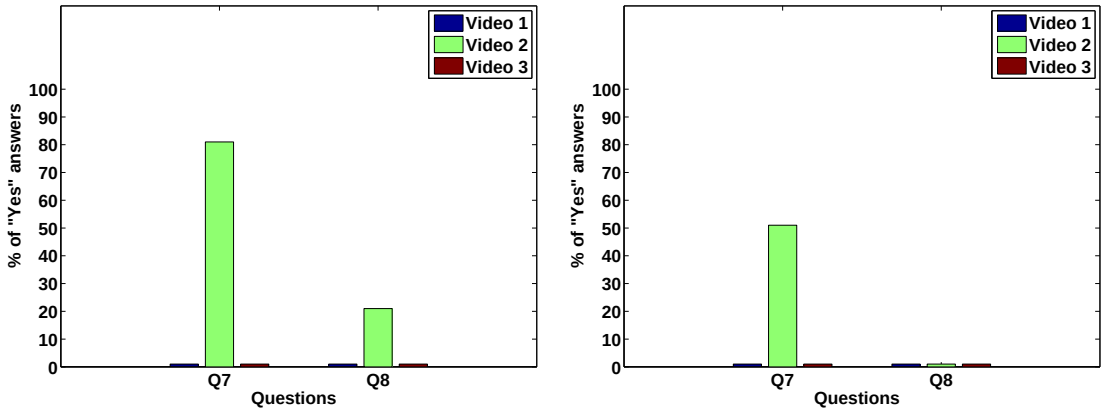
Analysis of *where* Evidence

As discussed earlier, the CCTV operator can use *where* information to infer identities. Figure 5.5 shows the results of *where* evidence related questions. While we removed the obvious sources of the *where* information (i.e. text legends and university logo), the local users could still recognize the place without much difficulty. Few users could not recognize particular premises (question Q6) as the video was recorded inside a lab. The local user could recognize the place because the users were familiar with the places of recording. Note that generally local CCTV operators



(a) Local Participants (b) Remote Participants

Figure 5.5: User study results for questions 3, 4, 5, and 6.



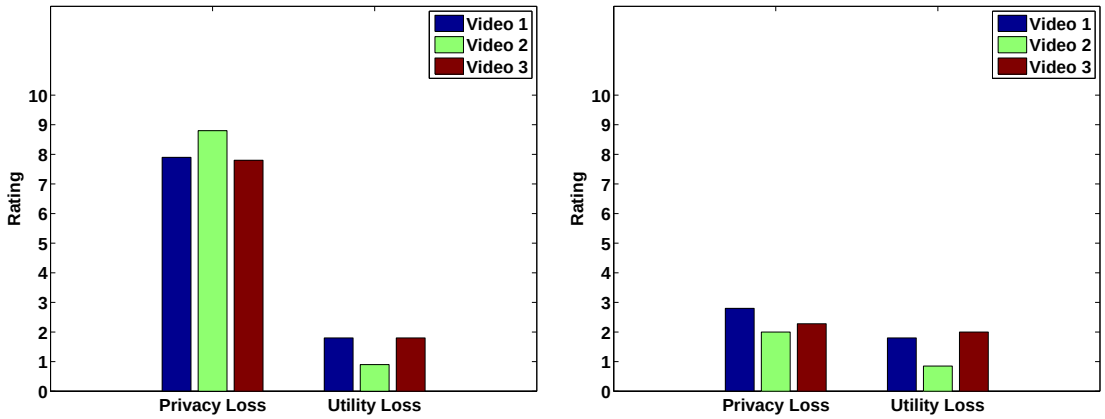
(a) Local Participants (b) Remote Participants

Figure 5.6: User study results for questions 7 and 8.

also have a good familiarity with the surveillance site. On the other hand, the remote users could only recognize the continent, but they could not recognize country, city, or premises.

Analysis of *when* Evidence

Figure 5.6 shows the results of *when* evidence related questions. The results show that at the local site also only a few users could recognize the time correctly. Although the local users were not able to determine the time precisely, in a real-time surveillance system they could always find time through external means like a watch or mobile phone. The remote users were unable to detect the time, and since they could not recognize the location, they were not able to use other means to get the time information. Hence, the remote viewing also helped in hiding the timing information.



(a) Local Participants (b) Remote Participants

Figure 5.7: Overall ratings of the users.

Overall Effectiveness

The local users rated the videos as almost full privacy loss, although they accepted that with given video clips they could easily perform the surveillance tasks. The remote users felt less privacy loss, while utility loss ratings are the same as the local users. Figure 5.7 shows the overall response of users with respect to the privacy loss and the utility loss.

Based on the analysis above and the results of the user study, we conclude that the current privacy protection methods do not apply to the traditional surveillance systems, and the system architecture itself need to be modified in order to provide true privacy. Hence, we propose anonymous surveillance system, where context decoupling is achieved by showing the video only to the people who do not have contextual overlap with the surveillance site where the context can have multiple dimensions of temporal, spatial, and cultural etc.

5.3 Anonymous Surveillance Framework

In traditional CCTV surveillance systems, the operators watch the camera feeds locally. The operators either already know the surveillance site in advance or get familiar after a while. In both cases, the CCTV operator gains enough contextual knowledge that s/he can infer the identity of individuals in the video. Even when *who, when* and *where* information is removed from the video using detectors (that are still not fully accurate leaving some faces visible), the operator can infer the site due to familiarity with place, can obtain current time, and use the

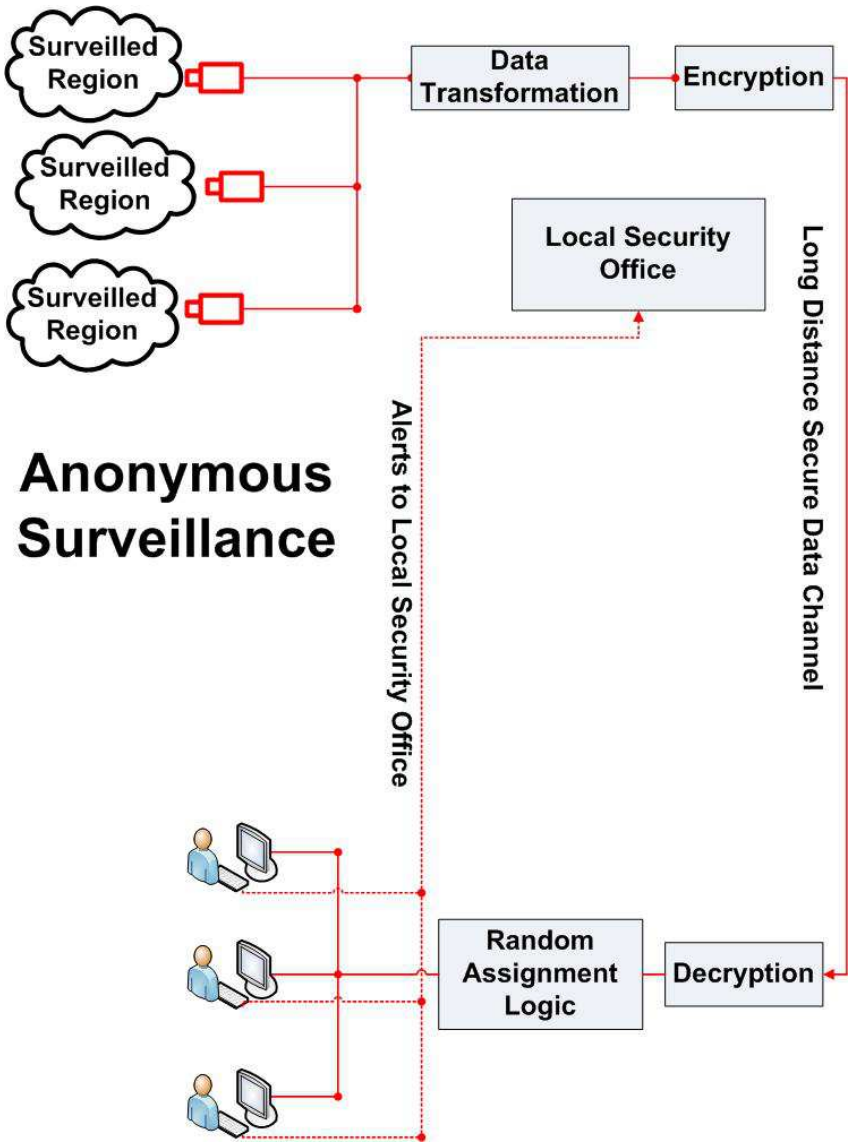


Figure 5.8: Anonymous Surveillance System. The black color is used to represent normal system components and red color is used to represent privacy-aware system components.

knowledge of usual patterns of activities to infer the identity.

In the proposed framework, the video feeds are shown to CCTV operator who are not familiar with the surveillance site so that removing text legends, logos, and other recognizable signs is enough for hiding *where* information. This is only possible if the video feeds are monitored remotely. Further, the remote place must be chosen such that it has very low probability of contextual overlap with the surveillance site. Here context can be defined over multiple dimensions i.e. geographical, temporal, social, etc. The remote monitoring also reduces the citizens feeling of privacy loss as recognized by Transportation Security Administration (TSA) while installing backscatter body scanners at airports [Kli08]. In this arrangement, the body scan is viewed at a remote location and the security officer at security check counter only receives the notification whether something is suspicious or normal. It is believed that not allowing security persons to see the body scan and the person being scanned simultaneously provides better privacy. At the end of this section we also arrive at similar conclusions for the privacy loss through surveillance cameras.

Figure 5.8 shows the overview of the proposed anonymous surveillance framework. Video camera feeds that are to be monitored remotely are first transformed to remove the *when* and *where* information and then sent over a long distance network to a distant place (probably another continent) after encryption. At the remote security office, these streams are decrypted and randomly assigned to the CCTV operators. The operators at the remote office are given instructions regarding the normal and abnormal situations, although no information about the monitored site is provided. In case of any abnormal situation, the viewers can anonymously signal the local security office. Consequently, the security personnel at the local office may access the surveillance site and take appropriate actions in real time immediately. In what follows, we provide the details of the framework and discuss the challenges involved.

5.3.1 Local Security Office

The security personnel working in this office do not have access to the camera feeds; however, they have knowledge of the locations of the cameras. In case they receive any signal of suspicious information from the remote office, they can go and manually inspect the area and take appropriate actions; or in extremely critical situations they can be provided access to the video

to make an educated assessment of the situation. There are security threats that can only be determined by fusing information from multiple cameras [AKJ06] e.g. emergency evacuation and stampede. These threats are detected at the local security office by analyzing the information received from the remote site and fusing it with local contextual knowledge.

The fusion of information over multiple cameras is carried out at the local site because the security personnel working at the remote site are only provided with limited context knowledge. This raises concern of what knowledge the remote group should be provided and how much information they need to send to the local office so that the fusion can be performed satisfactorily. Also, it needs to be analyzed how this will affect the privacy loss.

5.3.2 Data Transformation

The main motivation of remote surveillance is to remove adversary's knowledge related to the surveilled site. Therefore, it is important that we remove the location information from the video. In surveillance systems the cameras are generally static, which makes it easier to remove the location information. The text legends can be blurred or pixelized until they become meaningless [CNIB08]. Removal of popular figures and structures is more tricky and it might need shape deformation.

In the proposed framework, large amounts of the video data is transmitted over the Internet with limited bandwidth. Therefore, the data transformation method should be fast and it should be bandwidth friendly so that the size of the transformed data is minimized. In Section 5.4 we highlight the importance of the background information by modeling the privacy loss that occurs through background, and subsequently provide an evaluation of background anonymization methods from processing time and transformed data size perspective.

5.3.3 Data Protection

The sensitive video data travels over the Internet from the local site to the remote site. During this time, the data may travel over many untrusted networks; therefore, a suitable data protection scheme should be applied before sending data e.g. data encryption [CKM08]. More robust methods of data security are being proposed [WR11] that can be incorporated into the system for better reliability. Similarly, the data needs to be decrypted at the remote site for viewing

the content.

5.3.4 Camera Assignment

The proposed framework recommends a large cloud of CCTV operators monitoring cameras from all over the world, similar to the outsourced call centers. Every operator is asked to monitor a specific number of cameras. If an operator watches the same camera video for a long time, it might lead to leakage of *where* information through too much familiarity with the place. The operator can search the web or describe the place to others to get the location information. Once the location information is known, the adversary can establish a link between the remote site and local site leading to the privacy loss. To combat this situation, we choose to do random assignment of the cameras after a fixed quantum of time.

The periodic re-assignment of the cameras also increases the effectiveness of the CCTV operators by reducing dullness and fatigue that comes by monitoring the same video over long time. However, the cameras should be assigned such that every CCTV operator gets to observe almost equal number of targets. As the amount of attention (or work) required from an operator is proportional to the number of the targets in the camera view, we mention these targets as workload and propose a workload equalizing camera assignment method in Section 5.5.

5.3.5 Remote Security Office

All the CCTV cameras are monitored here. The viewers are not provided any information about the location of the camera or current time at the surveillance site. They are given a dictionary of rules which defines basic normal and abnormal situations at the site being monitored. The rules are kept anonymous so that no *when, where* information can be inferred from the rules. For example, instead of saying "The monitored site is a lobby of XYZ airport where heavy bags are not allowed" the viewer is told "Heavy bags are not allowed at the site being monitored". Other type of security threats like intrusion, weapons, fighting, theft, left baggage etc. generally do not need any *when* and *where* information.

As mentioned earlier, there are three main issues related to the remote site: (1) what is the minimal set of rules that need to be provided to the security personnel at the remote site for satisfactory quality of surveillance and high probability of privacy protection? (2) how does the

Table 5.3: Scenarios for user study #2

Scenario	Description
Scenario1	A CCTV operator is watching the video in the same building.
Scenario2	The CCTV operator is watching the video in a different country, but S/he knows the location of the camera.
Scenario3	The CCTV operator is watching the video in a different country and S/he does not know the location of the camera.
Scenario4	The CCTV operator is watching the video in a different country and S/he does not know the location of the camera. Further, after certain period of time the, the CCTV operator is changed.

security agent at remote site anonymously inform local security personnel about the current situation? and (3) what information is to be sent in normal situations and how will it change if the situation is not normal?

While precisely answering these questions is outside the scope of this thesis, we have performed a user study based experiment through which we found that the anonymous surveillance can provide both security and privacy. The details of the experiment are provided next.

5.3.6 User Study #2

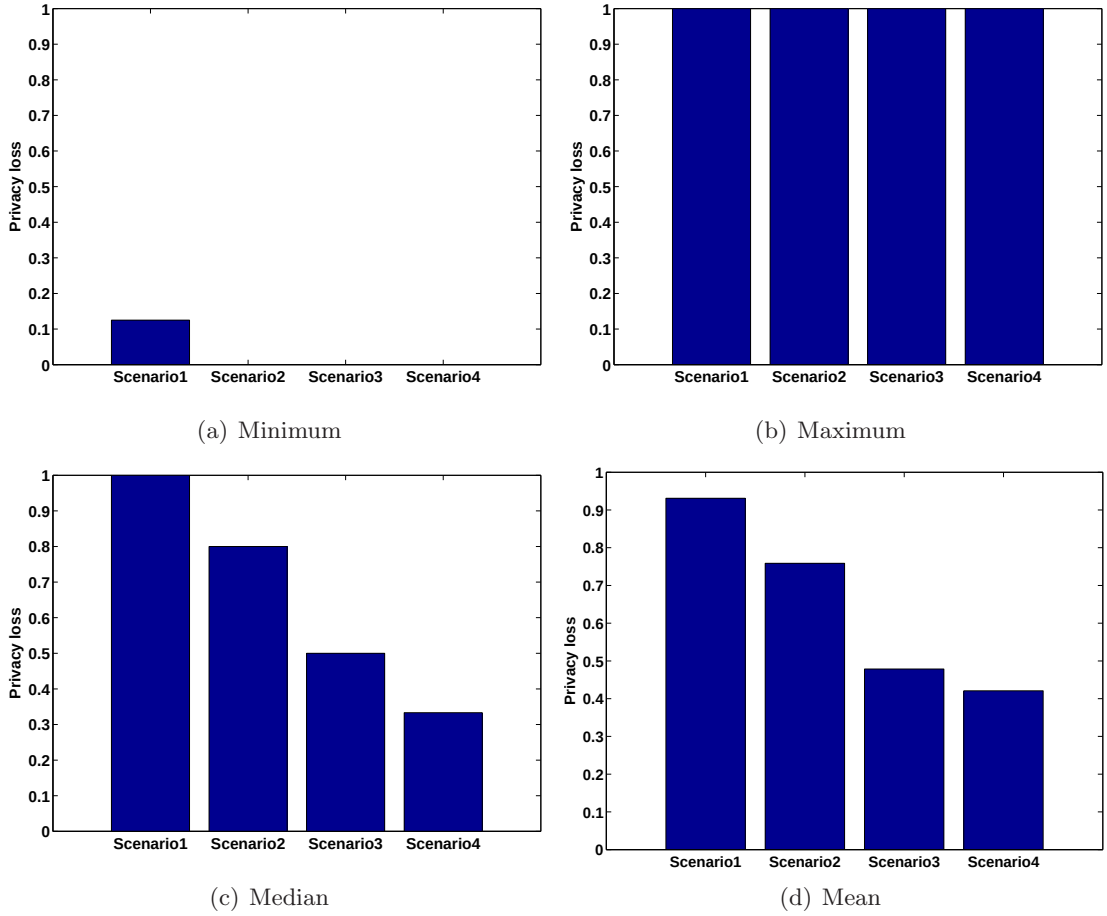
The end goal of the proposed framework is to provide a feeling of privacy protection to the surveilled people. To evaluate the proposed framework from this perspective, we conducted a user study. In this study we recorded opinion of over 100 participants from 18 countries spread across the globe. The users were asked to rate their feeling of privacy loss on a scale of 0 to 10 with 0 referring to no privacy loss and 10 referring to full privacy loss, in the four scenarios described in the Table 5.3.

For the comparison purpose, we normalized the ratings between 0 to 1 by dividing every users ratings by his/her maximum rating i.e. if $\mathcal{R}_i$ is the set of ratings provided by i^{th} user, the normalized ratings $\mathcal{R}_i^s$ are calculated as follows:

$$\mathcal{R}_i^s = \frac{\mathcal{R}_i}{max(R_i)} \tag{5.1}$$

The normalized user ratings are plotted in the Figure 5.9. We can draw the following conclusions from the results of the user study:

- The first bar is higher than the remaining three bars and the privacy loss decreases significantly from scenario 1 to scenario 2 which shows that users feel less privacy loss in anonymous surveillance. They get the feeling that the operator monitoring at the remote



(a) Minimum

(b) Maximum

(c) Median

(d) Mean

Figure 5.9: Results of the user study for privacy loss corresponding in four scenarios given in Table 5.3.

Table 5.4: Description of the video clips used for background anonymization

Video	Scenario	duration(m)
Clip 1	Corridor	45
Clip 2	Lab	40
Clip 3	Entrance	35
Clip 4	Canteen	45

place would not be able to recognize them.

- The feeling of privacy loss further decreases when the location is anonymized. The decrease in privacy occurs because users become more confident that the CCTV would not be able to identify them as s/he does not know the place. This supports our claim that users feel less privacy loss if their identity is preserved.

- In scenario 4, no operator is allowed to watch the same camera feed beyond a specified time. Users feel less privacy loss with this setting. This is in accordance to our observation that random assignment of the operators after certain quantum of the time would prohibit the viewer from building a context and eventually it would provide better sense of the privacy preservation.

From the observations of the user study we conclude that background anonymization and random assignment of the cameras are important for the privacy protection. Therefore, in next two sections we propose a bandwidth friendly background anonymization method and workload equalizing camera-to-operator assignment.

5.4 Background Anonymization

In the data transformation model of the Figure 5.8, we need to anonymize the background to hide the location information from the remote operators. Background anonymization requires additional processing of video which may cause timing overhead. The size of the transformed data is also important in a network based application. Further, any given anonymization method should cause minimal distortion to the video data. We evaluate the following three anonymization methods from the above mentioned perspectives: blurring, replacing by a black box, and pixelization. The representative frames from the videos used for this experiment are shown Figure 5.10 and the description of these videos is give in Table 5.4.

Processing times for the three methods are shown in Figure 5.11. From the figure it is

(a) Clip 1 (b) Clip 2

(c) Clip 3 (d) Clip 4

Figure 5.10: Representative background frames (after anonymization) from four video clips.

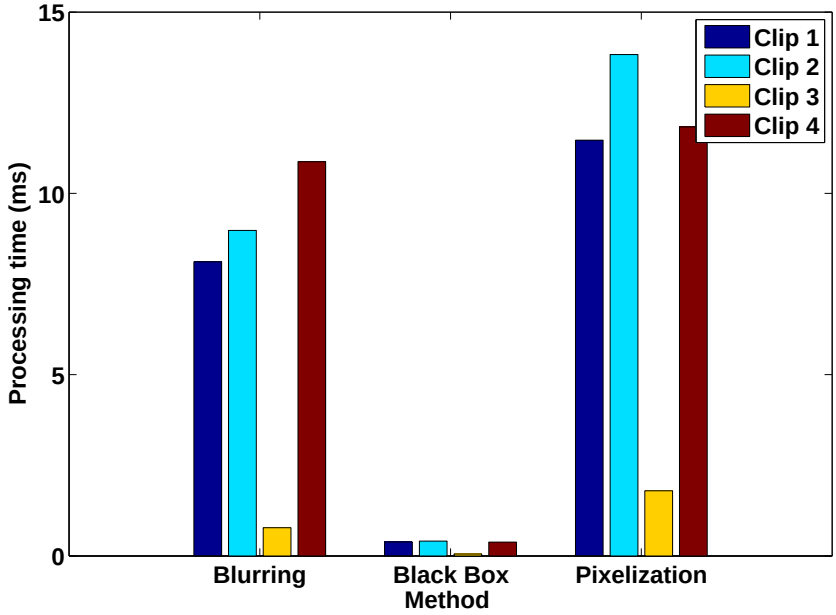


Figure 5.11: Average processing time per frame for background anonymization methods.

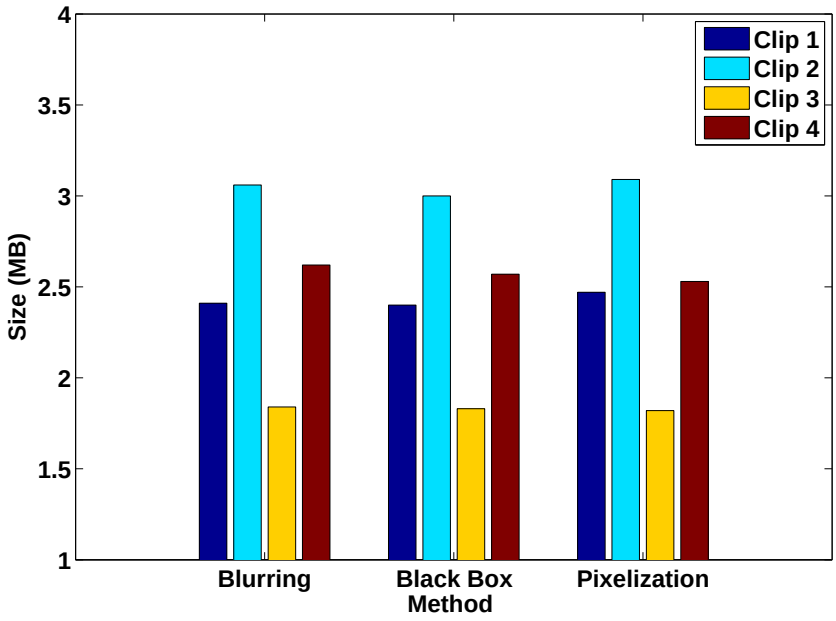


Figure 5.12: Average size of the transformed data.

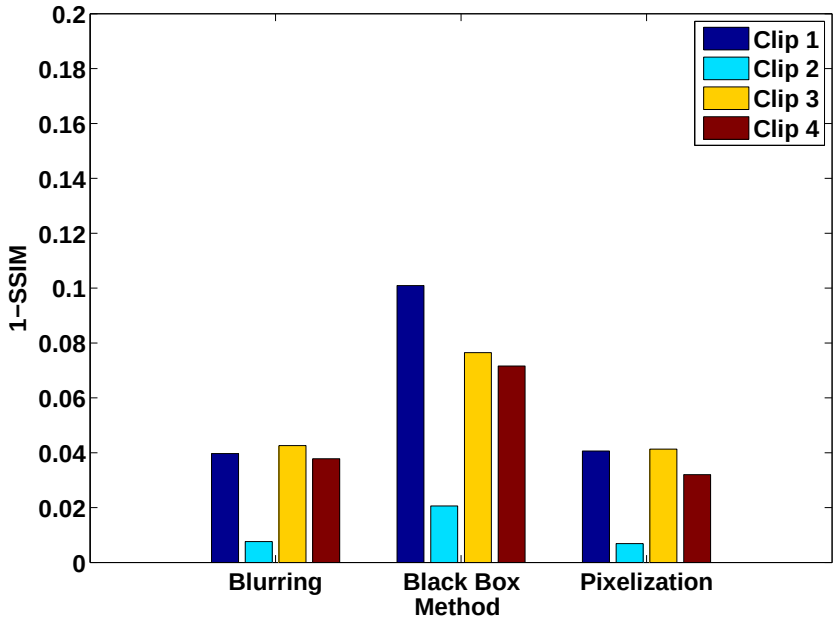


Figure 5.13: Average distortion measure (1-SSIM) for the three methods of background anonymization.

clear that replacement by a black box takes minimal time. This is logical as there is minimal processing required. The size of the transformed data is shown in Figure 5.12. We see that pixelization produces smaller sized data than blurring. In many applications the perceptual quality of the data becomes more important. Therefore, we calculate structural similarity index (SSIM) [WBSS04] to measure the perceptual quality of the transformed data in comparison to the original data. The distortion is measured as complement of SSIM. In Figure 5.13 we can see that pixelization produces the least distorted image while replacement by black box distorts the data maximally.

5.5 Random Assignment of Cameras to Remote Operators

There are mainly three motivations of random camera assignment in anonymous surveillance system: (1) reducing privacy loss (2) reducing fatigue of the operator and (3) equalizing dynamically changing workload among operators. By workload we mean the number of targets being monitored, which keeps changing over time for a given camera. Each target (people, vehicle etc.) in the camera view needs to be detected, recognized, tracked and analyzed to identify potential threats, which requires lot of attention of the CCTV operator. Therefore, it is necessary to distribute the cameras such that each operator has to monitor equal number of targets. This is in contrast to the situation when one operator in monitoring feeds with multiple targets in all frames and another operator is receiving video with no targets. Hence, static camera assignment does not allow for efficient resource utilization.

In anonymous surveillance system, video is transmitted over network rather than dedicated cables. In these settings, it is very easy to forward any video feed to any operator dynamically. We take this opportunity to explore a dynamic workload sharing method to equalize the amount of workload handled by each operator. The workload equalization is also important in the emerging paradigm of cloud based sensing infrastructures where the goal would be to maximize the utilization of the alloted processing power. Dynamic camera-to-operator assignment requires an appropriate model which captures the variability of the workload. For cost effective and efficient load sharing, we need to know the characteristics of the workload generated by each camera.

We propose a Markov chain based model of workload for video surveillance systems. Different

states of the Markov chain meticulously capture the semantics of the workload in terms of the number of targets. The monitoring load depends on the environmental conditions and it is found to be proportional to the number of targets. For example, more attention is required to process the video from a camera which is placed at a crowded place, than a camera with only few targets in its view. The model is validated with real surveillance data. Subsequently, we describe a dynamic load sharing method for load equalization among operators. .

5.5.1 Previous Work

In the past, researchers have proposed various workload models in different contexts. Maxiaguine et al. [MKT04] proposed *wcet* (Worst Case Execution Time) and *bcet* (Best Case Execution Time) to characterize the workload for real time embedded systems. The model is appropriate for understanding extreme behavior, yet, it does not capture the dynamic characteristics of the workload and is hard to use for load sharing. A Markov chain based model has been proposed by Song et al. [SEY05] which preserves the dynamic behavior of the workload in different states. The work mainly focuses on the parallel computers where the states are determined based on the number of nodes requested by the task. However, the model does not accommodate the semantics of the task. In our case the task of video monitoring depends on the number of targets. In contrast to the works described above, our model preserves the timing characteristics of the workload in the states of a Markov chain.

5.5.2 Workload Model

Figure 5.14 shows a topological view of a typical surveillance system installation. The media streams are generated at sensors and transmitted to the operators over the network where they are monitored to accomplish the surveillance goal. In the proposed workload model, we use a Markov chain to represent the number of targets (usually people and/or vehicles) in the environment. We define a target flow graph that can be easily constructed by observing the operating scenario. This graph is used to measure the transition probabilities and the steady state probabilities of the states. These states have been identified such that the workload shows similar behavior in each state.

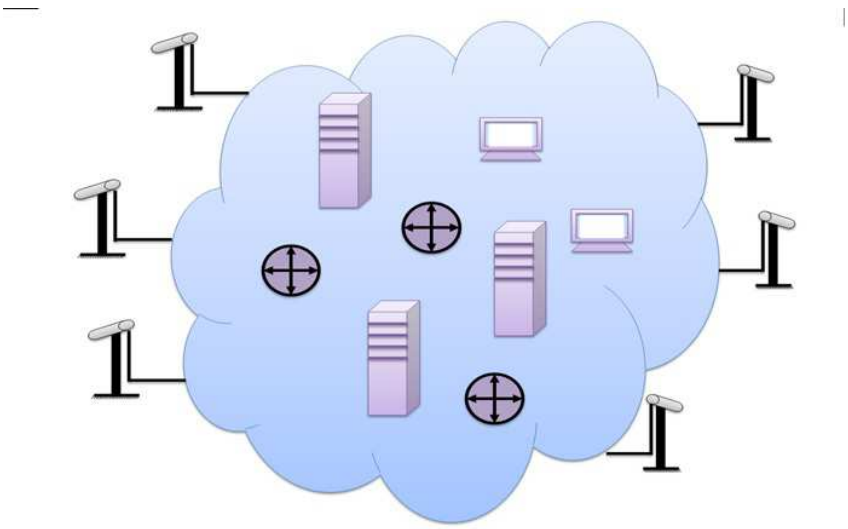


Figure 5.14: The cloud represents the network. The processing units and the users are distributed over the network.

Target Flow Graph (TFG)

Although surveillance systems are generally designed to observe human behavior, many times they need to monitor other type of objects as well, like abandoned baggage [SQGP06], vehicles [PGM10], etc. Hence, for clarity, we will use the term targets to describe humans and other objects, unless mentioned otherwise. Once the camera placement is fixed, we need to learn the dynamics of the environment. Actual workload is derived based on this knowledge and is represented as a target flow graph. The target flow graph TFG is constructed as a set of tuples:

$$TFG = \{(t_k, g_k) \mid k \in [1, l]\} \tag{5.2}$$

where g_k is the number of targets at t_k^{th} time instant and l is the total number of observations. In other words, we represent the flow of targets in a time series format i.e. number of targets in the coverage area at sampling instances. Figure 5.15 shows the target flow graph for a specific surveillance scenario.

Markov Chain Construction

A typical surveillance system usually needs to track each target individually and store the tracking results with other contextual information for activity analysis. As discussed earlier,

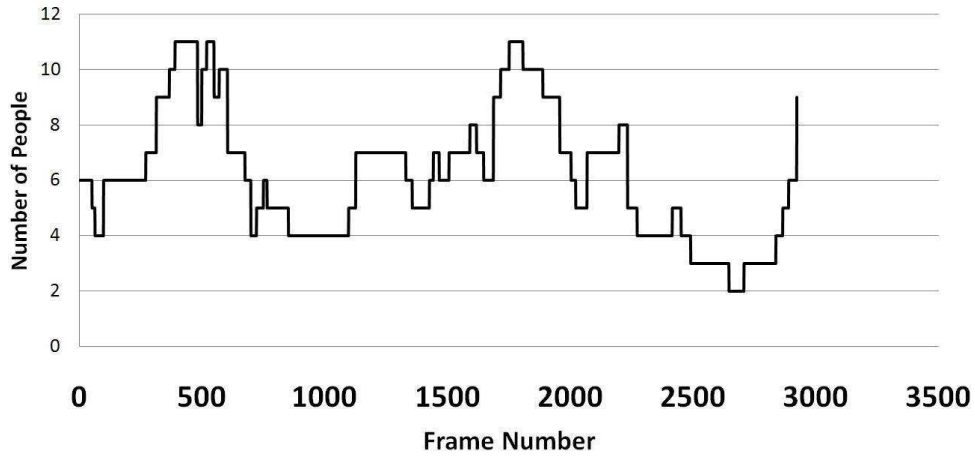


Figure 5.15: Target flow graph for a surveillance scenario.

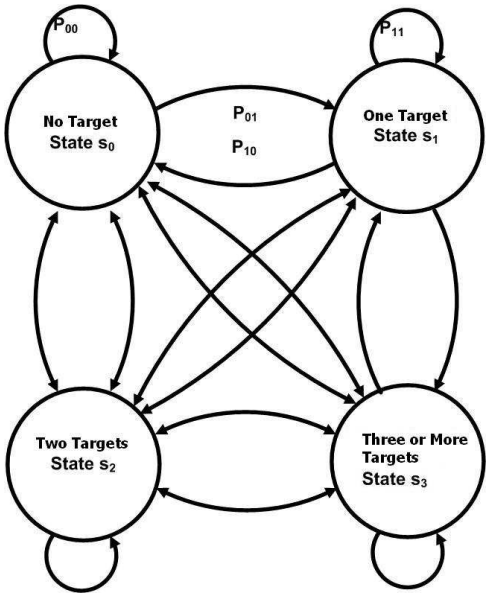


Figure 5.16: Different states of the Markov chain depending on the number of targets and transition probabilities.

unlike other domains, in surveillance the amount of attention for each frame mainly depend on the number of targets in the camera view. We capture this dependence in a Markov chain. We model the workload as a Markov chain because it can preserve the temporal behavior of the workload in its states and thus can capture the variability of the workload. The number of states in a Markov chain is $m + 1$ where m is given by the following equation:

$$m = max\{g_k \mid (t_k, g_k) \in TFG, k \in [1, l]\} \tag{5.3}$$

In other words, m is the maximum number of targets expected in the monitoring area. It is easy to get such information from analysis of the environment where the surveillance camera is installed. To capture the variability of the workload, we define different states of a Markov chain according to the number of targets. The set of states $\mathcal{Z}$ of this Markov chain can be defined as follows:

$$\mathcal{Z} = \{z_0, z_1, ...z_m \mid \forall i, j \in [0, m], z_i = i, z_i \neq z_j\} \tag{5.4}$$

Figure 5.16 illustrates the states of the Markov chain for a surveillance scenario. There are two types of probabilities associated with a Markov chain: transition probability and steady state probability. The transition probabilities are represented in the form of a matrix χ. The elements of the transition matrix can be constructed as follows:

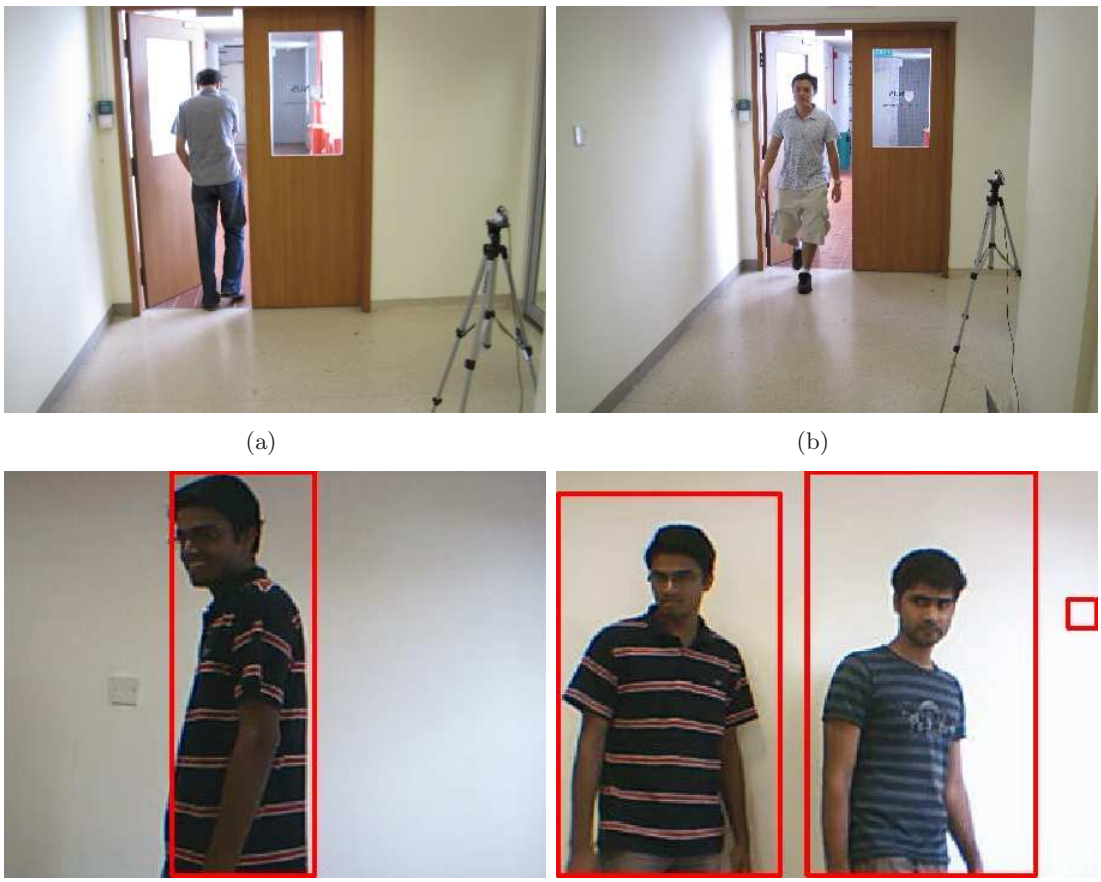
$$\chi = \{p_{ij} \mid p_{ij} = \frac{n_{ij}}{n_i}, i, j \in [0, m]\} \tag{5.5}$$

where n_i is the number of times the camera is in state z_i and n_{ij} is the number of times the camera transiting from state z_i to state z_j, and these are calculated as

$$n_i = \mid \{g_k = i, (t_k, g_k) \in TFG, k \in [1, l]\} \mid \tag{5.6}$$

$$n_{ij} = \mid \{g_k = i \wedge g_{k+1} = j, (t_k, g_k) \in TFG, k \in [1, l-1]\} \mid \tag{5.7}$$

The transition probabilities capture the dynamic nature of the workload. These probabilities, along with the steady state probabilities, can be used to predict the future workload given the



(a) (b)

(c) (d)

Figure 5.17: System Installation.

current state of the workload. Let $\Pi = (\pi_0, \pi_1, \pi_2, ...\pi_m)$ be the steady state probabilities of the states. These can be calculated using one of the following two ways:

$$(\chi - I)\Pi = 0 \tag{5.8}$$

$$\Pi = \{\pi_i \mid \pi_i = p'_{ij}, p'_{ij} \in \chi^{Inf}, i = x, j \in [0, m]\} \tag{5.9}$$

where I is identity matrix, Inf is a large number (≈ 100), and x is any number between 0 and m. In fact, all the rows of the matrix χ^{Inf} will have similar values. The equation (8) gives the eigenvector of the transition probability matrix for the eigenvalue of 1.

Table 5.5: The specifications of the system.

Operating System	Microsoft Windows XP
Platform	Visual C++ 2008
Additional Libraries	OpenCV
Computer	Intel(R) T2300 @ 2.33GHz, 0.99GB Ram
Image Resolution	320 × 240 captured by AXIS IP camera

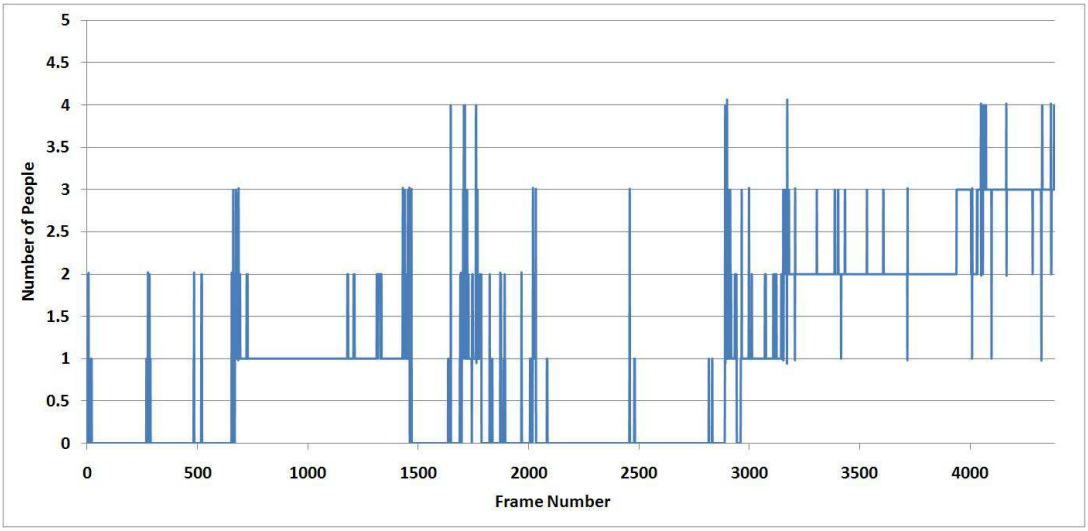


Figure 5.18: Target flow graph for the implemented system.

Model Validation

Because measuring human attention is hard, we validate the model by implementing an automated surveillance system that does similar tasks, although with very low accuracy. The system is first trained to learn the background; anything that appears in the foreground is considered as a target. When there are multiple targets, the system tracks each of them individually until they disappear from the camera view. Whenever targets appear in the camera view, a new object is created. When the target leaves the camera view, the object is deleted and the tracking history is stored to disk. Table 5.5 lists the hardware and software details of the implemented system.

The experiments were conducted in a corridor near the exit which is a typical surveillance setting (Figure 5.17). As the first step of modeling, we record the target flow pattern and construct the TFG as shown in Figure 5.18. The TFG is then used to calculate the transition

Table 5.6: The state-wise values of mean and variance of processing times for blob detection and tracking.

State/Targets	μ(milliseconds)	σ(milliseconds)
0	496	24
1	518	30
2	557	30
3	662	38
4	733	78

and the steady state probabilities. The transition probability matrix is given by:

$$\chi = \begin{pmatrix} 0.9814 & 0.0150 & 0.0026 & 0.0005 & 0.0005 \\ 0.0214 & 0.9023 & 0.0595 & 0.0149 & 0.0019 \\ 0.0077 & 0.0702 & 0.8904 & 0.0285 & 0.0033 \\ 0.0091 & 0.0183 & 0.0639 & 0.8858 & 0.0228 \\ 0.0435 & 0.1739 & 0.1304 & 0.3043 & 0.3478 \end{pmatrix} \tag{5.10}$$

and steady state probability vector is given below:

$$\Pi = \begin{pmatrix} 0.4344 & 0.2473 & 0.2107 & 0.1020 & 0.0057 \end{pmatrix} \tag{5.11}$$

Similarly, we collect time statistics to calculate mean and variance of processing time. The experimental results are given in Table 5.6. It is apparent in the table that the processing time is proportional to the number of targets.

5.5.3 Dynamic Load Sharing

The workload model proposed in the previous section is used for dynamically assigning video streams to operators. From the workload model we understand that the amount of resources required depends on the state of the environment being observed by the camera. These states can be dynamically calculated for each camera. Now, if there are N_{pc} operators and $C_{pc}(j)$ is the set of cameras assigned to j^{th} operator, our objective is to find an assignment scheme which maximizes the equalization function:

$$E_{pc} = \frac{1}{\kappa} \prod_{j=1}^{N_{pc}} \sum_{\forall k; c_k \in C_{pc}(j)} z(c_k) \tag{5.12}$$

$$\kappa = \left(\frac{1}{N_{pc}} \sum_{j=1}^{N_{pc}} \sum_{\forall k; c_k \in C_{pc}(j)} z(c_k) \right)^{N_{pc}} \tag{5.13}$$

where c_k is the k^{th} camera, $z(c_k)$ is the state of that camera and κ is a normalization coefficient. By state of the camera we mean the state of the environment being monitored by the camera. If the state variables are always unity, this would result in every operator tracking equal number of targets. However, different cameras can be in different states resulting in non uniform workload. The transition probabilities of the states predict future behavior of the workload. For example, if a camera reaches a very high state but its probability of transitioning to lower states is high, it might be more beneficial to retain the old camera assignment as this is a transient state which will diminish quickly. Note that changing camera assignment also has associated cost that can be minimized by using transition probabilities.

Although the attention required increases with the number of targets, a minimal amount of vigilance is still required when no targets are detected (Table 5.6). Therefore, we keep the number of the cameras for each operator fixed while manipulating the camera-to-operator assignment. Following are the main steps of the dynamic load sharing method:

1. Perform the re-assignment task every w seconds. A larger value of w would result in lower number of re-assignment but it will also reduce E_{pc}. It can be assumed that there is a trusted entity that controls the camera assignment.

2. For each operator, check the total number of targets it is processing and calculate the average number of targets per operator (L_{av}).

$$L(j) = \sum_{\forall k; c_k \in C_{pc}(j)} z(c_k) \tag{5.14}$$

$$L_{av} = \frac{1}{N_{pc}} \sum_{j=1}^{N_{pc}} L(j) \tag{5.15}$$

   The number of targets can be easily calculated using blob detection methods. The detection task can be done on the high quality data at the local security office and the target information can be transmitted to remote office as video metadata.

3. Divide the operators into two groups: (1) $ListH' = (L(j) > L_{av})$ (2) $ListL' = (L(j) <$

L_{av}). The first list contained the overloaded operators whereas the second list contains the workload deficient operators.

4. Sort the lists according to the number of targets: $ListH = SORT(ListH')$ ans $ListL = SORT(ListL')$.

5. Perform the re-assignment in decreasing order of workload in $ListH$ and increasing order in $ListL$ until one of the lists is empty. The re-assignment of cameras between operators $OP1 \in ListH$ and $OP2 \in ListL$ is done in the following steps:

   (a) Pick the cameras from $OP1$ that have higher workload than average ($L^c_{av} = L_{av}/N^c_{av}$) and that are likely to face higher workload in future. If the sum of the state probabilities of the states higher than L^c_{av} is more than a threshold ($i.e. PROB(s(c) > L^c_{av}) > P_{th}$), it is likely that they will face high workload in the future. These are the cameras facing high workload and need to be re-assigned to other operators.

   (b) Similarly, pick the cameras from $OP2$ which have workload lower than average and which are likely to face lower workload in the future. If the sum of the state probabilities of the states lower than L^c_{av} is higher than a threshold ($i.e. PROB(s(c) < L^c_{av}) > P_{th}$), it is likely that they will face low workload in the future. These cameras can be assigned to the operators facing high workload in exchange to the camera in high state.

   (c) Swap the picked camera's monitoring tasks between the operators $OP1$ and $OP2$.

   (d) Repeat these steps until cameras on all of the operators are all checked or we achieve workload equalization (i.e. $L(j) \leq L_{av}$).

5.5.4 Experiments and results

In the experiments we demonstrate the advantage of the proposed load sharing method. We simulate a distributed surveillance system with 100 cameras and 20 operators ($N_{pc} = 20$). All operators are assumed to be of equal capability. We keep the number of cameras assigned to the operators fixed to five and vary their assignment to operators in re-assignment phase. The value of w is taken to be one. The probability threshold P_{th} for camera selection is 0.5 as the probability below this would mean favoring undesirable transitions. We use two performance

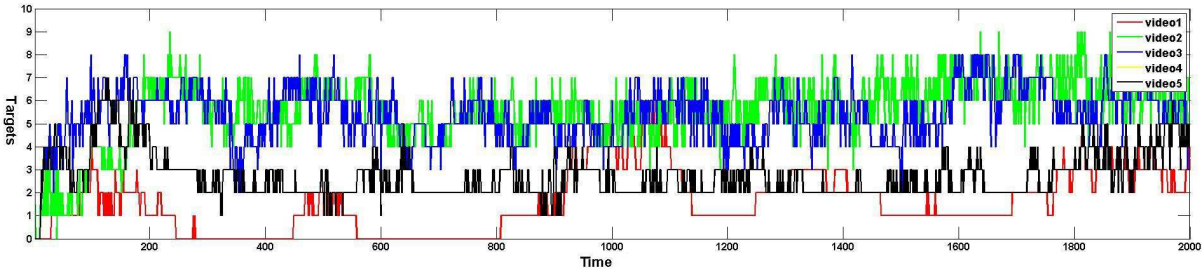


Figure 5.19: The target flow graph of the five scenarios corresponding to the PETS [PET11] videos.

measures to evaluate our methods: (1) Equalization (E_{pc}) (2) Number of targets dropped. Since five cameras are assigned to a operator, and there are 10 states in our Markov model, we assume that a operator can handle 25 targets simultaneously at a time. The targets in excess to 25 are considered dropped.

The experiments are divided into three parts. In the first experiment we show the effect of different static camera assignment schemes on the number of targets dropped and the E_{pc} (Section 5.5.4). The second experiment evaluates the proposed dynamic load assignment method with respect to the static assignment (Section 5.5.4). Finally in the third experiment, we show how transition probability helps in reducing the number of camera re-assignments (Section 5.5.4).

Dataset

Five different videos from PETS [PET11] are used to simulate five surveillance scenarios. Each of these videos consists 2000 frames taken at 2000 time instants. We extract the blob information from these videos and simulate a distributed system in Matlab to evaluate the performance of the proposed method. It is assumed that every scenario has 20 cameras. The data for 20 cameras is obtained using the same video but shifting the time axis and copying. First we shift each video by 50 frames to create 10 videos and then copy these 10 videos to create 20 videos. We copy the videos to simulate the scenarios where two cameras are placed to monitor same site from two different angles. Time shifting simulates the case when the cameras are placed at adjacent sites. The target flow graph corresponding to all five videos is shown in Figure 5.19. These target flow graphs are used to calculate the state transition probabilities in Equation 5.16 - 5.20.

$$\chi_1 = \begin{pmatrix} 0.9896 & 0.0104 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.0063 & 0.9399 & 0.0522 & 0.0016 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.0801 & 0.8811 & 0.0388 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.0544 & 0.9014 & 0.0442 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.1171 & 0.8108 & 0.0721 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.1194 & 0.8507 & 0.0299 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1.0000 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \tag{5.16}$$

$$\chi_2 = \begin{pmatrix} 0.75 & 0.25 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.0263 & 0.5 & 0.3421 & 0.0789 & 0.0526 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.3235 & 0.4118 & 0.2647 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.1091 & 0.0909 & 0.5636 & 0.1818 & 0.0182 & 0.0364 & 0 & 0 & 0 \\ 0.005 & 0 & 0.005 & 0.0452 & 0.4523 & 0.3367 & 0.1307 & 0.0251 & 0 & 0 \\ 0 & 0 & 0 & 0.002 & 0.1402 & 0.4248 & 0.3293 & 0.0894 & 0.0102 & 0.0041 \\ 0 & 0 & 0 & 0.0015 & 0.038 & 0.2398 & 0.4898 & 0.1901 & 0.0409 & 0 \\ 0 & 0 & 0 & 0.0026 & 0.0052 & 0.112 & 0.3516 & 0.4115 & 0.1042 & 0.013 \\ 0 & 0 & 0 & 0 & 0 & 0.0842 & 0.2632 & 0.4316 & 0.2 & 0.0211 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.6 & 0.3 & 0.1 \end{pmatrix} \tag{5.17}$$

$$\chi_3 = \begin{pmatrix} 0.9 & 0.1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.5 & 0.5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.1111 & 0.5556 & 0.3333 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.0667 & 0.4 & 0.4667 & 0.05 & 0.0167 & 0 & 0 & 0 \\ 0 & 0 & 0.0083 & 0.0661 & 0.595 & 0.2837 & 0.0413 & 0.0055 & 0 & 0 \\ 0 & 0 & 0 & 0.0095 & 0.1556 & 0.573 & 0.2333 & 0.0286 & 0 & 0 \\ 0 & 0 & 0 & 0.0032 & 0.0288 & 0.2336 & 0.6064 & 0.1184 & 0.0096 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.0632 & 0.3043 & 0.5455 & 0.087 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.0213 & 0.1277 & 0.4468 & 0.4043 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \tag{5.18}$$

$$\chi_4 = \begin{pmatrix} 0.9 & 0.1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.2083 & 0.625 & 0.1667 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.015 & 0.847 & 0.134 & 0.004 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.0046 & 0.1973 & 0.6995 & 0.0895 & 0.0091 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.0421 & 0.2947 & 0.5263 & 0.1211 & 0.0158 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.0513 & 0.3333 & 0.5256 & 0.0897 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.0571 & 0.2 & 0.6571 & 0.0857 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.3333 & 0.6667 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \tag{5.19}$$

$$\chi_5 = \begin{pmatrix} 0.9 & 0.1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.25 & 0.75 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.0143 & 0.8571 & 0.1286 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.0545 & 0.8622 & 0.0769 & 0.0064 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.0539 & 0.8529 & 0.0858 & 0.0074 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.004 & 0.0656 & 0.8032 & 0.1193 & 0.008 & 0 & 0 \\ 0 & 0 & 0 & 0.0028 & 0.0085 & 0.1676 & 0.7415 & 0.0739 & 0.0057 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.0155 & 0.1503 & 0.7668 & 0.0674 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.1974 & 0.7895 & 0.0132 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix} \tag{5.20}$$

Experiment #1: Effect of Static Camera Assignment

Generally cameras are permanently assigned to the operator. In this experiment we randomly pick two static camera assignments and calculate the E_{pc} and number of targets dropped. The results are shown in Figure 5.20. It can be observed that in some regions of the graph, assignment1 gives better performance whereas in some regions assignment2 gives better results. Hence the static assignment of the cameras is undesirable and we need dynamic workload sharing method.

Experiment #2: Workload Equalization using Dynamic Load Assignment

In this experiment we use the proposed method to dynamically share the workload among the operators. The re-assignment is performed at each time instant. Figure 5.21a compares the results of the proposed method with static assignment. In the static assignment, the five

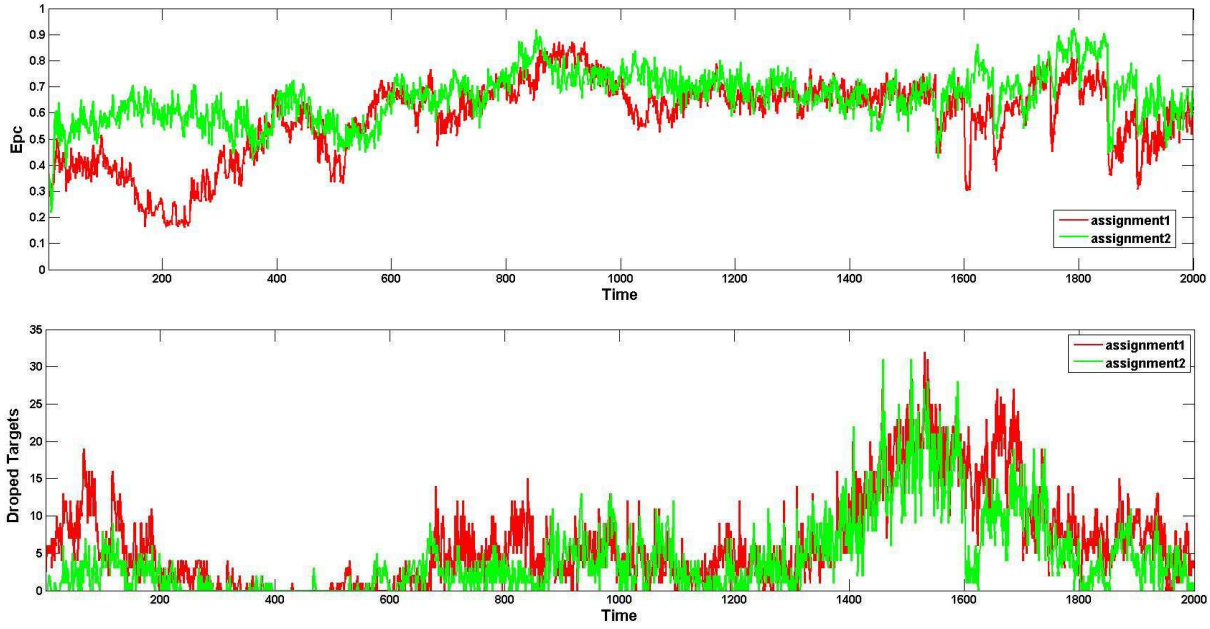
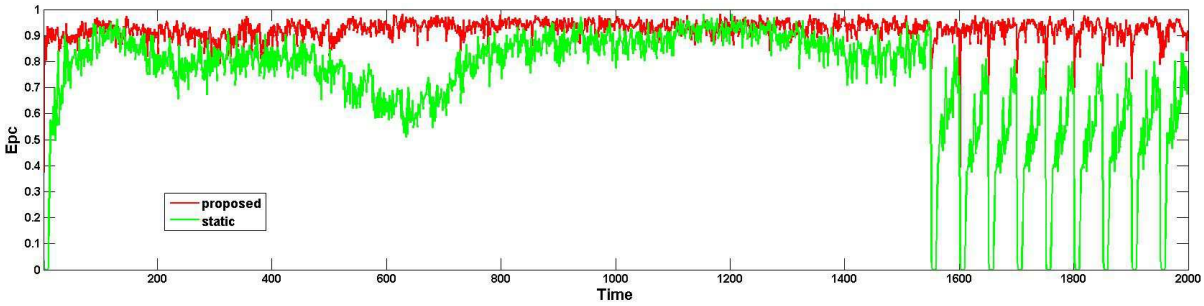


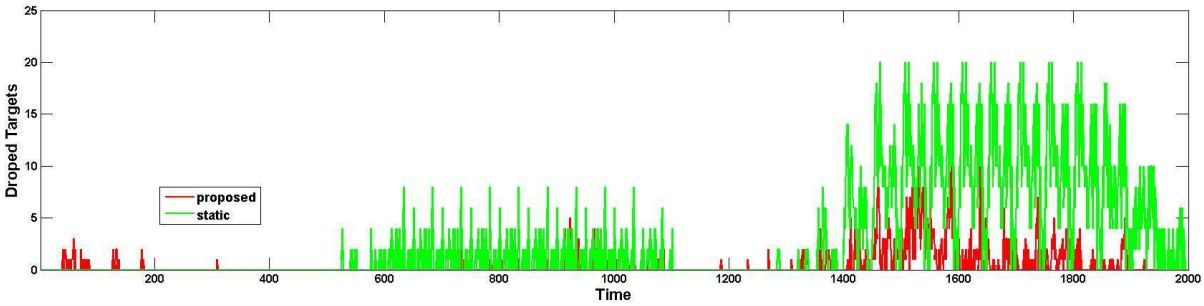
Figure 5.20: E_{pc} and number of targets dropped for two random static camera assignments.

cameras assigned to a operator are taken from five different scenarios. It can be seen that the proposed method almost always achieves better equalization compared to the static assignment. The E_{pc} for static assignment sometimes becomes zero. These are time instants when at least one of the operators is having zero targets to be processed. Here we want to clarify that Figure 5.19 only shows target flow graph for five cameras. Other cameras will have target flow graph which is shifted on time-axis.

To calculate the number of targets dropped, we assume that every operator can only monitor a limited number of targets, which is a realistic assumption. After every re-assignment we calculate the number of dropped targets. The results are shown in Figure 5.21b. In these results, the proposed method performs better than the static method. However, at some time instants (mainly between time instants 0 to 200) the proposed method drops more targets than static method. This happens because of the probabilistic nature of our method as the prediction may not always be correct. However, on average the proposed method has significant improvement over static method by providing 44% better equalization and reducing the target drop rate by 83%.



(a) Equalization values for static and proposed method.
Average value for proposed method =0.922 and static method =0.729



(b) Dropped targets for static and proposed method.
Average value for proposed method =902 and static method =5360

Figure 5.21: Dynamic vs. static load assignment results: (a) Workload equalization (E_{pc})(b) Number of targets dropped.

Table 5.7: Effect of Transition probabilities.

	States	States and Transition
E_{pc}	0.918	0.922
Targets Dropped	955	902
Re-assignments	6664	6445

Experiment #3: Overhead due to Dynamic Load Assignment

The camera re-assignment after equalization introduces overhead in terms of information transfer from one operator to another operator. It is desirable to achieve maximum equalization with minimum the number of re-assignments. In this experiment we perform equalization using two methods. In one method we only use the current state information (i.e. the number of targets at that instant) whereas in another experiment we perform equalization using transition probabilities as described in proposed method. The results are shown in Table 5.7. We observe that with the help of transition probabilities we are able to reduce the number of targets dropped by 6% and number of re-assignments by 3.2%. This shows that transition probabilities reduce the re-assignment overhead on the system while maintaining the quality.

5.5.5 Discussions

The CCTV operators need to monitor each target individually in order to do activity and event detection. We propose a target based workload model for surveillance systems which is found to be close to real workload. Based on this workload model, a dynamic load sharing method is proposed to do workload equalizing camera-to-operator assignment. The experimental results show that the proposed dynamic workload sharing method performs better than traditional static workload assignment. It is also found that the transition probabilities of the Markov model help in reducing the overhead of camera re-assignment. The proposed method can be used in the random camera assignment phase of the anonymous surveillance system. Additional constraints need to be imposed to make sure that after re-assignment each operator watches different cameras than earlier.

5.6 Conclusions & Summary

It is very difficult to provide robust privacy preservation in traditional surveillance systems. The CCTV operator has strong prior knowledge of the surveillance site and the habitants, which can

cause significant privacy loss even in the absence of the bodily cues. It is found that in order to reduce this privacy loss, the context knowledge of the CCTV operator needs to be decoupled from the surveillance site. Hence, we propose anonymous surveillance framework where the surveillance camera are monitored remotely by operators who are kept unaware of the location of the camera.

We further identified that in anonymous surveillance systems, background anonymization and workload equalizing camera assignment are two important problems. An anonymity based model is proposed to measure the privacy loss that occurs due to background information even when no people are present in the video. To the best of our knowledge, this is the first attempt to model privacy of absence. For equalized assignment of the cameras such that each operator monitors equal number of targets, a Markov chain based model of the workload is proposed. This model is used to build a dynamic workload assignment scheme. The experiments show that the proposed methods is able to obtain 44% better equalization with 83% less number of targets dropped in comparison to earlier static assignment based methods.

Chapter 6

Summary, Conclusions and Future Work

6.1 Summary

In this dissertation we have looked at the problem of privacy loss in video surveillance systems from a novel perspective. Contrary to the earlier works which only consider the bodily cues of the identity leakage, we identify that the privacy loss could occur even when the facial information is hidden. In this work we analyze the privacy breach that exists in the current privacy protection methods in the field of privacy-aware video surveillance. As our first finding, we recognize that the privacy loss is a function of the identity leakage and the sensitive information that is present in the video. Both of these factors are necessary for an adversary to gain sensitive information about the individuals. Therefore we measure the privacy loss as a product of identity leakage and sensitivity index (a measure of the sensitive information in the video). While the sensitive information mainly depends on the user priorities, the identity leakage is determined based on the adversaries knowledge.

In Chapter 3, we propose a novel model of the privacy loss from the public access of the video data of a single camera. The identity leakage of the individuals in the video is calculated based on the presence of *who, what, when* and *where* information. It is found that the adversary can use this information to infer identities even without facial information using his/her prior knowledge. The proposed model calculates the identity leakage as the certainty of the adversary regarding

the presence of an individual in the camera view. Subsequently, we propose a hybrid method to transform the video data such that no individuals can be identified. The transformed data can be published for various research purposes. We also propose a task based model to measure the utility of the published data. Through experiments we show that for the applications with blob detection and tracking as main tasks, first blurring the data and then followed by quantization produces the best tradeoff between the utility and the privacy.

The model described above is sufficient to measure the privacy loss from a single camera video. However, in Chapter 4 we show that when the adversary gets access to video from multiple cameras, additional identity leakage can occur through correlation among events and activities from different cameras. To measure the privacy loss in a multiple camera scenario, we extend the privacy model to explicitly include the adversary's knowledge. We model the adversarial knowledge as set of propositions and propose an enhanced anonymity based framework for calculating the identity leakage of the individuals in the video. The identity leakage is calculated as inverse of the anonymity of each individual. The additional inference channels due to multi-camera video are highlighted in the experiments.

The main goal of this work is to analyze and improve the robustness of the privacy protection methods in privacy-aware surveillance systems. The findings of the privacy loss through implicit identity leakage channels lead us to a very important conclusion that it is very difficult to provide robust privacy in traditional surveillance systems. The adversary (CCTV operator in this case) has contextual knowledge of *when* and *where* which cannot be removed through computer vision techniques. Based on this finding, we propose an anonymous surveillance framework in Chapter 5 that decouples the adversary's knowledge from the video data through remote monitoring. In the user study we find that the proposed framework provides robust privacy preservation by blocking *when* and *where* information. In the proposed anonymous surveillance framework, the video is transmitted over network to the operators. Because it is easy to switch videos in these scenarios, we proposed a Markov chain based workload model for surveillance task and workload equalizing camera-to-operator assignment method.

6.2 Contributions

The main contributions of thesis are on the privacy modeling and its application in the privacy-aware surveillance and data publication as follows:

1. The existing works consider only explicit identity leakage (mainly facial information), but they do not taken into account the implicit channels of *what, when* and *where*. To the best of our knowledge, this is the first attempt to model the privacy loss as a continuous variable considering both explicit and implicit channels.

2. Most of the earlier works have modeled the privacy loss as the identity leakage alone or the presence of sensitive information alone. However, the privacy loss is function of both of these quantities. In this thesis we quantify the identity leakage and the sensitivity index and propose a model that combines these quantities to calculate overall privacy loss.

3. We model the utility loss for the application of data publication and propose a hybrid data transformation method (using a combination of quantization and blurring). This provides an opportunity of publishing surveillance video data which can be very useful for testing vision algorithms, video ethnography, and policy making.

4. In traditional surveillance systems, the CCTV operator has prior context knowledge about the surveillance site and its habitants, which makes it difficult to block the implicit identity leakage channels. We propose an anonymous surveillance framework that advocates decoupling the contextual knowledge from the video. The experiments show that the proposed framework is effective in blocking the identity leakage channels and provides better sense of the privacy to the individuals.

5. The surveillance task is target (people, vehicle, etc.) centric and the amount of attention depends on the number of the targets in the camera view. We use a Markov chain to model this workload and propose a dynamic workload assignment method that equalizes the number of targets monitored by each operator by dynamically changing the camera-to-operator assignment.

6. Privacy loss could still occur when no people are present in the camera view through background information. We call it absence privacy and propose model to quantify it.

Subsequently, a network friendly background anonymization method is determined based on the evaluation with respect to size of transformed data, processing time, and visual distortion.

6.3 Conclusions

The following are the conclusions from the research work carried out in this thesis:

- Current privacy protection methods only provide false sense of privacy protection; there is a privacy breach that exists due to *what, when* and *where* information present in the video. This privacy breach has not been considered by the previous works.
- The hybrid method of first blurring and then quantization provides the best utility vs privacy tradeoff in case of data publication application for blob detection and tracking tasks.
- When an adversary has access to multiple camera video, the identity leakage of the individuals further increases. The adversary can track the individuals over multiple cameras, observe the places they visit, and infer the identity based on this information.
- The background information can cause privacy loss even when no individuals are present in the video. For background anonymization, pixelization produces transformed data of smaller size and takes minimum processing time when compared to blurring and quantization.
- The dynamic assignment of the camera to the operators equalizes the workload among CCTV operators to increase the surveillance effectiveness and reduce the operator fatigue.

In this thesis we have proposed various methods to measure the privacy loss and the utility loss. Our findings from this thesis lead us to the a number of research problems that need to be solved for a robust privacy preservation.

6.4 Future Research Directions

In this thesis we have highlighted the privacy breach in the current privacy preservation techniques and consequently propose methods that minimize the utility loss and provide robust

privacy. The work can be extended in multiple directions as described below. While first two sections describe our immediate future works, remaining sections summarize potential future research directions.

6.4.1 Trajectory Anonymization for Video Data Publication

Video collections are growing in size due to pervasive use and dropping cost of multimedia devices. For example, large number of cameras are being deployed in surveillance systems to increase the coverage area [DV08]. The video recorded from these cameras has large amount of information that is very useful for many applications. However, publishing video data may cause privacy loss of the individuals if the viewers are able to identify them in the video. The video needs to be anonymized before publication so that no the identity information can be inferred from it. Traditional method of preserving privacy in the video is to hide facial information. However, when the adversary has access to video recorded at multiple location, s/he can learn the trajectory of the individuals in the video. This trajectory information can lead to the identity leakage [NAS08]. In this work we want to anonymize the trajectory to block the identity inference with minimal compromise in the utility of the data.

The location information from different videos can enable the adversary to learn the trajectories of the individuals. It has been found in the previous works that the trajectory information can reveal the identity [NAS08]. Many works have been reported on the trajectory anonymization for point trajectories (where trajectory is published as set of location-time tuples). While anonymizing a point trajectory data is easy, it is hard to anonymize the video shots that form the trajectory. It is very hard to modify two video trajectories such that they look similar to the viewer. Even the location information can either be present or absent, but it cannot be partially removed to make multiple trajectories appear similar.

Related Works

To the best of our knowledge, there are no works on video trajectory anonymization. Nonetheless, current methods of point trajectory anonymization are the following:

- The initial works on trajectory anonymization exploited the uncertainty of the GPS system in determining the exact location of the users. All the trajectories in a cluster are shifted in

spatial domain to lie within the uncertainty region [ABN08]. In another work [HGXA07], some trajectory points are removed to increase the uncertainty between consecutive points.

- There are suppression based techniques to preserve the identity inference from the trajectories. The authors in [TM08] suppress a trajectory such that an adversary cannot infer the tail of a trajectory given the head of the trajectory. The authors argue that when an adversary has knowledge of partial trajectory of a person, s/he can use this knowledge as a quasi identifier and infer the remaining locations the person visited. Therefore, they propose to remove location information from certain point in the trajectory.

- The current trend in trajectory anonymization is to bring the concept of k-anonymity and cluster the trajectories such that each cluster has at least a specified number of trajectories [NAS08]. The challenge in this technique is to define appropriate distance measure.

- In other works, some portions of the trajectory are manually labeled as quasi identifiers (QIDS) by user and subsequently anonymized [BWJ05]. The challenge here is to identify which portions of the trajectory are quasi-identifiers. Based on our initial observation, the beginning and the end parts of the trajectory might be good candidates for quasi-identifiers. Yet, any part of the trajectory can be identity revealing. For example the adversary may have the knowledge that only Mukesh takes a particular path and infer the identity.

Problem Formulation

A video from multiple cameras may have trajectory information embedded in it. The adversary can learn this trajectory and use this information to infer the identities. Therefore, the first step would be to detect the trajectories in a given set of videos. The trajectory of each individual can be represented as a sequence of shots (from different cameras). Once the trajectories are detected successfully, the video is transformed to perturb these trajectories so that they do not cause any identity leakage.

Following are the other scenarios in which a video can have trajectories:

- A moving camera records the trajectory as an individual moves around in the space. In this scenario, even when the person is replaced by a blob, the trajectory can be determined.

However, the trajectory length would be generally limited to a small premise.

- The video is recorded from a static camera showing multiple locations e.g. an omni-direction camera mounted on the ceiling of a corridor. This video can capture the trajectory of individuals although the length would be limited to the camera view.

- The video is recorded by multiple static cameras, each camera representing one 'point' of the trajectory. The video shots from these cameras are put together to form a trajectory. However, in its original form, the trajectory can cause identity leakage. Therefore need to anonymize the trajectory such that the identity leakage does not happen.

Irrespective of the source of the video, for modeling purposes we assume that we have a video that consists of multiple shots corresponding to the trajectory of an individual in the video. We want to find a transformation function $\mathcal{T}$ that minimizes both the utility loss U and the privacy loss Γ for the given video V i.e.

$$E = \eta\Gamma(\mathcal{T}(V)) + (1-\eta)U(V, \mathcal{T}(V)) \tag{6.1}$$

where η is a normalizing coefficient. The main challenges in trajectory anonymization problem involves modeling of the privacy loss Γ and the utility U; and then finding appropriate transformation function T that minimizes E.

Applications

We need to derive a utility model after fixing the set of application/queries that will use the published trajectory data. The usual applications of the trajectory data are:

- In data mining for example traffic planning, marketing.

- By governments in understanding the infrastructure [Pen99].

- In social sciences to study human behavior [CR06] [LCMA07], .

- By Companies to improve the efficiency of the employees [MO06].

Preliminary Solutions

If we have a video clip that is recorded at several locations, it will have trajectory information. Video shot detection methods can be used to determine portions of the video that correspond to different locations. Now the following transformations can be used to reduce the privacy loss from these videos:

- **Information Hiding** Video is reduced in the quality such that people, locations etc. cannot be determined. For example, the video can be blurred, pixelized, and/or quantized to the extent that people and locations are unidentifiable.
- **Shot Dropping** Identify the sensitive portions of the video and drop them. For example, the initial and final parts of the trajectory generally enable the adversary to learn the identity, therefore these can be dropped.
- **Shot Reshuffling** Once the video shots are detected, these shots can be reshuffled so that the adversary cannot learn the trajectory information.
- **Video Fabrication** This is very interesting method of the identity preservation. In this we insert staged video shots into the original video such that the shots from the original video are anonymized.

The tradeoff between the privacy and the utility can be modeled in many ways, such as maximizing the number of frames which have the motion information in the resulting video; or minimizing the number of spurious frames added to the video in order to do the anonymization.

6.4.2 Motion Similarity Index (MSIM) for Evaluating Data Transformation Methods

In the privacy preserving surveillance, the video is transformed to such a form that the viewer cannot identify the individuals. Many methods, such as blurring, pixelization, black box, scrambling etc. have been proposed for this purpose. The performance of these methods is measured in terms of the robustness with which they provide privacy; however, the effect of data transformation on the surveillance is generally overlooked. The quality of the transformed video is measured based on the visual distortion calculated for each image e.g. PSNR and SSIM

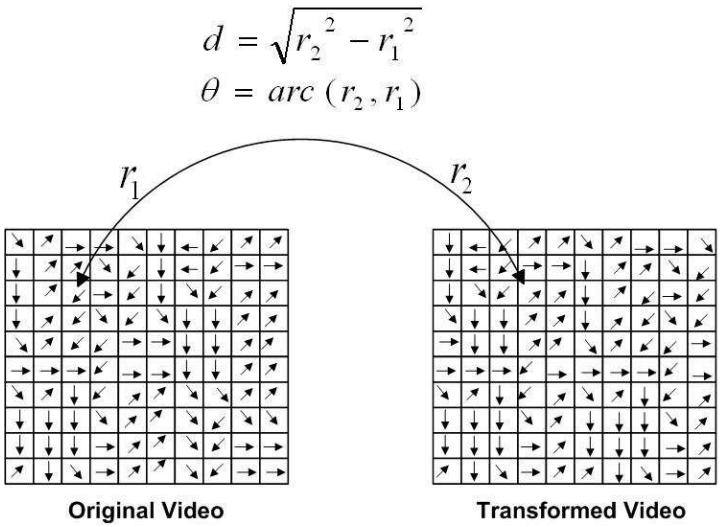


Figure 6.1: The comparison of motion between two images of a video.

[WBSS04]. However, most of the surveillance tasks depend more on the ability to recognize the activities in the video rather than perceptual quality. Therefore, the motion distortion of the video should be calculated and fused with the visual distortion to calculate overall distortion index of the anonymized videos.

We recognize that simple perceptual or structural distortion measures of video quality are not appropriate for comparing the surveillance quality of the video. This is because these performance measures only consider the static properties of individual images but fail to consider the motion information of the video. Note that the motion is the most desired characteristics of a surveillance video. In surveillance scenarios, where the main task is to identify suspicious activity or behavior detection, the motion plays an important role. If motion of the video is preserved, it can be sufficient for surveillance tasks even when it does not have other visual aesthetics.

In this work, we want to model a motion distortion based quality index for comparing two videos with respect to surveillance quality. In this method we compare both the angle and distance of the motion vectors of original and transformed video and obtain a single quality index between zero and one (Figure 6.1). In the second part of the work, we want to compare various data transformation techniques.

Preliminary Method

We first calculate the visual distortion using the method described in [WBSS04]. Then we calculate the motion distortion between the videos. Finally we combine both the measures to calculate the overall index. There are many challenges in comparing the motion of two videos and deriving a single index to represent the whole video. Following steps can be taken to compare the motion of two videos and calculate the distortion in the transformed video:

- Calculate the block based motion vectors in both the images.
- Calculate the difference in the motion vectors, and normalize them between zero and one using some method.
- Calculate this motion difference for all corresponding blocks as shown in the Figure 6.1.
- Obtain a cumulative motion distortion index for the frame.
- To get the cumulative motion distortion between the videos, calculate the weighted sum of the motion distortion of the frames.
- The weights are calculated based on the amount of the motion they have. A frame with more amount of motion gets more weight. A frame with no motion gets zero weights. This approach will filter out all the static frames which do not have any surveillance related information in them.

Validation

We need to show that the proposed model is more accurate in providing the quality assessment of the video in comparison to earlier methods. Particularly we can compare three performance measures: (1) PSNR (2) SSIM (3) MSIM (proposed method). There are mainly two steps involved in this: (1) transform the video using different methods (2) and compare the quality measures with the surveillance operator perceived quality.

6.4.3 Adversary Knowledge Modeling

Weather the information in the video will cause privacy loss or not depends heavily on the adversary's knowledge. The privacy models take certain assumptions in order to measure the privacy

loss. For example, in the proposed multi-camera privacy model, the adversarial knowledge is approximated in the propositional statements which are later used to measure the identity leakage and the privacy loss. Therefore an accurate model of the adversary's knowledge is very useful and it can provide more precise privacy loss. We intend to explore the following issues related to adversary knowledge modeling.

- How to measure the adversary's knowledge? The adversary has access to the voter list, news papers, and news channels. All these sources need to be considered in order to calculate the privacy loss of the individuals. If we take analogy of the relational data publication, it is assumed that the published data can be matched with already available data and leak the identity. This makes it very important to determine what is already *available* to the adversary.

- In the proposed privacy model, the adversary's knowledge is modeled as set of propositions. However, the adversarial knowledge changes over time as s/he gains access to more video data. It is desirable to accurately measure the adversary's knowledge for the privacy loss measurement. Therefore, in future we want to develop methods to automatically populate the adversarial knowledge and update dynamically with time.

- In the anonymous surveillance, minimal context knowledge is provided to the remote operator. Providing more contextual knowledge can increase chances of the privacy loss while less contextual knowledge can affect the surveillance quality. Therefore, we need to obtain an optimal sharing of the contextual knowledge between local and remote operator.

- The surveillance activity information is detected and transmitted to the local security office so that the local operators, who have full contextual knowledge, can fuse information from multiple remote operators and make better assessment of the situation. It needs to be determined what information should be sent from the remote operators to the local operators so that the information can reach in timely manner, is sufficient for surveillance, and does not provide any hidden channels of the identity leakage.

6.4.4 Data Transformation

While the privacy modeling is necessarily the first step, it is always followed by video data transformation. The data transformation should provide robust privacy with least compromise in the utility of the data. Following are the future research challenges in data transformation.

- Two contributing factors of privacy have been identified: identity leakage and sensitive information. The existing research work focuses towards hiding the identity of individuals. It would be interesting to explore removal of the sensitive information from the video. The main problem here is to determine what information is sensitive and what is suspicious so that we only remove the sensitive information. It is difficult to automatically separate sensitive from suspicious information because both of these have common characteristics of high entropy.

- As an alternative to the anonymous surveillance framework, important objects and activities from the camera can be mapped onto a virtual world. This transformation will have effect on privacy as well as quality of surveillance. To help surveillance application, some of the objects can be copied from the actual video, while the identity leaking regions can be anonymized. We need to find an optimal set of virtual and real image regions to construct the transformed video.

- All the current works assume manually detecting the events. These methods are not scalable over large amounts of the video. It is a challenge to make these methods scalable.

6.4.5 System Integration

The deployment of the privacy protection methods into real surveillance systems poses further challenges. We have identified following issues that arise in integrating the privacy protection methods into surveillance systems.

- Current event detection methods are not robust enough. Therefore, the proposed privacy preservation may fail when the vision algorithms fail. One solution to this problem is to take a conservative approach and use lower thresholds in detection algorithms so that we can get the privacy at the cost of increased false positives. In future we want to explore

how low thresholds are good for the privacy preservation and what is their effect on the utility.

- The proposed anonymous surveillance framework advocates remote monitoring for the purpose of context decoupling. However, such long distance networks are unreliable and cannot provide real-time guarantees. In future we want to implement a complete system and study its reliability.

These research challenges will lead us towards robust privacy protection in surveillance systems with minimal privacy loss. With the help of accurate privacy models, appropriate quality assessment measures, and data transformation functions that optimize the tradeoff between the utility and the privacy.

List of Publications

Book Chapters

Saini, M.; Atrey, P.; and Kankanhalli, M. *Workload Modeling for Multimedia Surveillance Systems*. In Emerging Paradigms in Machine Learning. Springer 2011.

Conferences and Workshops

[1] Saini, M.; Atrey, P.; Mehrotra, S.; and Kankanhalli, M. *Anonymous Surveillance*. In IEEE International Workshop on Advances in Automated Multimedia Surveillance for Public Safety with IEEE International Conference on Multimedia & Expo, Barcelona, 2011.

[2] Saini, M.; Wang, X.; Atrey, P.; and Kankanhalli, M. *Dynamic Workload Assignment in Video Surveillance Systems*. In IEEE International Conference on Multimedia and Expo. Barcelona, Spain. 2011.

[3] Saini, M.; Atrey, P.; Emmanuel, S.; and Kankanhalli, M. *Functionality Delegation in Distributed Surveillance Systems*. In IEEE International Workshop on Multimedia Systems for Surveillance in conjunction with IEEE International Conference on Advanced Video and Signal-Based Surveillance. Boston, USA. pp.72-79, 2010.

[4] Saini, M.; Atrey, P.; Mehrotra, S.; Emmanuel, S.; and Kankanhalli, M. *Privacy Modeling for Video Data Publication*. In IEEE International Conference on Multimedia and Expo. Singapore. pp.60-65, 2010.

[5] Saini, M.; Nataraj , Y.; and Kankanhalli, M. *Performance Modeling of Multimedia Surveil-*

lance Systems. In IEEE International Symposium on Multimedia, San Diego. pp.179-186, 2009.

[6] Saini, M.; Jain, R.; and Kankanhalli, M. *A Flexible Surveillance System Architecture.* In International Conference on Video and Signal Based Surveillance. Genova, Italy. pp.571-576, 2009.

[7] Saini, M.; Kankanhalli, M. *Context-Based Multimedia Sensor Selection Method.* In IEEE International Conference on Advanced Video and Signal Based Surveillance. Genova, Italy. pp.262-267, 2009.

[8] Saini, M.; Singh, V.; Jain, R.; and Kankanhalli, M. *Multimodal observation systems.* In ACM International Conference on Multimedia. Vancouver, British Columbia, Canada. pp.933-936, 2008.

[9] Mahapatra, D.; Saini, M.; Ying S. *Illumination invariant tracking in office environments using neurobiology-saliency based particle filter.* In IEEE International Conference on Multimedia and Expo. Hanover, Germany. pp.953-956, 2008.

Journals

[1] Saini, M.; Atrey, P.; Mehrotra, S.; and Kankanhalli, M. *Adaptive transformation for robust privacy protection in video surveillance.* Hindawi International Journal of Advances in Multimedia. (Accepted: February 6, 2012).

[2] Saini, M.; Wang, X.; Atrey, P.; and Kankanhalli, M. *Adaptive workload equalization in multi-camera surveillance systems.* IEEE Transaction on Multimedia. (Accepted: January 18, 2012).

Bibliography

[ABN08] O. Abul, F. Bonchi, and M. Nanni. Never walk alone: Uncertainty for anonymity in moving objects databases. In *IEEE International Conference on Data Engineering*, pages 376 –385, 2008.

[AKJ06] P. Atrey, M. Kankanhalli, and R. Jain. Information assimilation framework for event detection in multimedia surveillance systems. *Multimedia systems*, 12(3):239–253, 2006.

[AMCK+02] J. Al-Muhtadi, R. Campbell, A. Kapadia, M.D. Mickunas, and S. Yi. Routing through the mist: Privacy preserving communication in ubiquitous computing environments. In *IEEE International Conference on Distributed Computing Systems*, pages 74–83, 2002.

[AS95] M. S. Ackerman and B. Starr. Social activity indicators: interface components for cscw systems. In *ACM symposium on User interface and software technology*, pages 159–168, 1995.

[Ass48] U.N.G. Assembly. Universal declaration of human rights, 1948.

[BA07] P. Bhaskar and S. I. Ahamed. Privacy in pervasive computing and open issues. In *International Conference on Availability, Reliability and Security.*, pages 147 –154, 2007.

[BD99] D. Banisar and S. Davies. Privacy and human rights. In *www.gilc.org/privacy/survey*, 1999.

[BEG00] M. Boyle, C. Edwards, and S. Greenberg. The effects of filtered video on awareness and privacy. In *The ACM Conference on Computer Supported Cooperative Work*, pages 1–10, 2000.

[Ber00] A.M. Berger. Privacy mode for acquisition cameras and camcorders, 2000. US Patent 6,067,399.

[BKUT09] N. Babaguchi, T. Koshimizu, I. Umata, and T. Toriyama. Psychological Study for Designing Privacy Protected Video Surveillance System: PriSurv. *Protecting Privacy in Video Surveillance*, pages 147–164, 2009.

[Bou05] T.E. Boult. Pico: Privacy through invertible cryptographic obscuration. In *Computer Vision for Interactive and Intelligent Environment*, pages 27 – 38, 2005.

[Bra05] J. Brassil. Using mobile communications to assert privacy from video surveillance. In *IEEE International Parallel and Distributed Processing Symposium.*, page 8 pp., 2005.

[Bra09] J. Brassil. Technical challenges in location-aware video surveillance privacy. *Protecting Privacy in Video Surveillance*, pages 91–113, 2009.

[BS03] A.R. Beresford and F. Stajano. Location privacy in pervasive computing. *Pervasive Computing, IEEE*, 2(1):46 – 55, 2003.

[BWJ05] C. Bettini, X. Wang, and S. Jajodia. Protecting privacy against location-based personal identification. *Secure Data Management*, pages 185–199, 2005.

[BZK+90] S.A. Bagues, A. Zeidler, C. Klein, C.F. Valdivielso, and I.R. Matias. The right to privacy. *Harvard Law Rev.*, 4(5):193–200, 1890.

[BZK+10] S.A. Bagues, A. Zeidler, C. Klein, C.F. Valdivielso, and I.R. Matias. Enabling Personal Privacy for Pervasive Computing Environments. *Journal of Universal Computer Science*, 16(3):341–371, 2010.

[CAMN+03] R. Campbell, J. Al-Muhtadi, P. Naldurg, G. Sampemane, and M. D. Mickunas. Towards security and privacy for pervasive computing. In *International Conference on Software security: theories and systems*, pages 1–15, 2003.

[Cav07] A. Cavallaro. Privacy in video surveillance. *Signal Processing Magazine, IEEE*, 24(2):168 –166, 2007.

[CB07] A. Chattopadhyay and T.E. Boult. Privacycam: a privacy preserving camera using uclinux on the blackfin dsp. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1 –8, 2007.

[CCV07] J. Chaudhari, S. Cheung, and M.V. Venkatesh. Privacy protection for life-log video. In *IEEE Workshop on Signal Processing Applications for Public Security and Forensics.*, pages 1 –5, 2007.

[CCYY07] D. Chen, Y. Chang, R. Yan, and J. Yang. Tools for protecting the privacy of specific individuals in video. *EURASIP Journal on Applied Signal Processing*, 2007(1):107–107, 2007.

[Che11] S. Chesterman. *One Nation Under Surveillance: A New Social Contract to Defend Freedom Without Sacrificing Liberty (Introduction).* Oxford University Press, 2011.

[CKM08] P. Carrillo, H. Kalva, and S. Magliveras. Compression independent object encryption for ensuring privacy in video surveillance. In *IEEE International Conference on Multimedia and Expo*, pages 273–276, 2008.

[CNIB08] K. Chinomi, N. Nitta, Y. Ito, and N. Babaguchi. Prisurv: privacy protected video surveillance system using adaptive visual abstraction. In *Proceedings of the International Conference on Advances in multimedia modeling*, pages 144–154, 2008.

[CPN08] S. Cheung, J.K. Paruchuri, and T.P. Nguyen. Managing privacy data in pervasive camera networks. In *IEEE International Conference on Image Processing.*, pages 1676 –1679, 2008.

[CR06] F. Calabrese and C. Ratti. Real time rome. *Networks and Communication Studies*, 20(3-4):247–257, 2006.

[CVP+09] S. Cheung, M.V. Venkatesh, J.K. Paruchuri, J. Zhao, and T. Nguyen. Protecting and managing privacy information in video surveillance systems. In *Protecting Privacy in Video Surveillance*, pages 11–33. Springer London, 2009.

[CZT05] H.S. Cheng, D. Zhang, and J.G. Tan. Protection of privacy in pervasive computing environments. In *International Conference on Information Technology: Coding and Computing.*, volume 2, pages 242 – 247 Vol. 2, 2005.

[Dal77] T. Dalenius. Towards a methodology for statistical disclosure control. *Statistik Tidskrift*, 15(429-444):2–1, 1977.

[DB92] P. Dourish and S. Bly. Portholes: supporting awareness in a distributed work group. In *ACM SIGCHI conference on Human factors in computing systems*, pages 541–547, 1992.

[Duf11] F. Dufaux. Video scrambling for privacy protection in video surveillance: recent results and validation framework. In *Proceedings of SPIE*, 2011.

[DUM10] A. Dehghantanha, N.I. Udzir, and R. Mahmod. Towards a pervasive formal privacy language. In *IEEE International Conference on Advanced Information Networking and Applications Workshops*, pages 1085 –1091, 2010.

[DV08] H.M. Dee and S.A. Velastin. How close are we to solving the problem of automated visual surveillance? *Machine Vision and Applications*, 19(5):329–343, 2008.

[DvGN+08] A. Doulamis, L. van Gool, M. Nixon, T. Varvarigou, and N. Doulamis. First ACM international workshop on analysis and retrieval of events, actions and workflows in video streams. In *ACM International Conference on Multimedia*, pages 1147–1148, 2008.

[Dwo06] C. Dwork. Differential privacy. In *International Colloquium on Automata, Languages and Programming*, pages 1–12, 2006.

[FBRG11] C. Fernández, P. Baiget, F. X. Roca, and J. Gonzílez. Determining the best suited semantic events for cognitive surveillance. *Expert Systems with Applications*, 38(4):4068–4079, 2011.

[Fer10] D. Ferrucci. Build watson: an overview of deepqa for the jeopardy! challenge. In *International Conference on Parallel Architectures and Compilation Techniques*, pages 1–2, 2010.

[FKRR93] R.S. Fish, R.E. Kraut, R.W. Root, and R.E. Rice. Video as a technology for informal communication. *Communications of the ACM*, 36(1):48–61, 1993.

[FNT04] D.A. Fidaleo, H.A. Nguyen, and M. Trivedi. The networked sensor tapestry (nest): a privacy enhanced software architecture for interactive analysis of data in video-sensor networks. In *ACM International Workshop on Video Surveillance & Sensor Networks*, pages 46–53, 2004.

[FRVM+10] D. Freni, C. Ruiz Vicente, S. Mascetti, C. Bettini, and C.S. Jensen. Preserving location and absence privacy in geo-social networks. In *ACM International Conference on Information and knowledge management*, pages 309–318. ACM, 2010.

[FWCY10] B. Fung, K. Wang, R. Chen, and P. Yu. Privacy-preserving data publishing: A survey on recent developments. In *ACM Computing Surveys*, volume 42, 2010.

[HGXA07] B. Hoh, M. Gruteser, H. Xiong, and A. Alrabady. Preserving privacy in GPS traces via density-aware path cloaking. In *ACM Conference on Computer and Communications Security (CCS)*, 2007.

[HRWL84] F. Hayes-Roth, D. Waterman, and D. Lenat. *Building expert systems.* Addison-Wesley, Reading, MA, 1984.

[HS96] S. E. Hudson and I. Smith. Techniques for addressing fundamental privacy and disruption tradeoffs in awareness support systems. In *ACM conference on Computer supported cooperative work*, pages 248–257, 1996.

[KKH04] I. Kitahara, K. Kogure, and N. Hagita. Stealth vision for protecting privacy. In *International Conference on Proceedings of the Pattern Recognition*, pages 404–407, 2004.

[Kli08] D. Klitou. Backscatter body scanners-a strip search by other means. *Computer Law & Security Report*, 24(4):316–325, 2008.

[KTB06] T. Koshimizu, T. Toriyama, and N. Babaguchi. Factors on the sense of privacy in video surveillance. In *ACM Workshop on Continuous Archival and Retrival of Personal Experences*, pages 35–44, 2006.

[KvB90] K.P. Karmann and A. von Brandt. Moving object recognition using an adaptive background memory. *Time-varying image processing and moving object recognition*, 2:289–296, 1990.

[Lan01] M. Langheinrich. Privacy by design - principles of privacy-aware ubiquitous systems. In *International Conference on Ubiquitous Computing*, pages 273–291. Springer-Verlag, 2001.

[LCMA07] D. Luper, D. Cameron, J.A. Miller, and H.R. Arabnia. Spatial and temporal target association through semantic analysis and GPS data mining. In *International Conference on Information and Knowledge Engineering*, pages 25–28, 2007.

[Lev06] A. Levin. Is workplace surveillance legal in canada? In *ACM International Conference on Privacy, Security and Trust*, 2006.

[LGS97] A. Lee, A. Girgensohn, and K. Schlueter. Nynex portholes: initial user reactions and redesign implications. In *Proceedings of the international ACM SIGGROUP conference on Supporting group work: the integration challenge*, pages 385–394. ACM, 1997.

[LGW02] Y. Lu, W. Ga, and F. Wu. Automatic video segmentation using a novel background model. In *The IEEE International Symposium on Circuits and Systems*, pages 807–810, 2002.

[LHYK03] K.C. Lee, J. Ho, M.H. Yang, and D. Kriegman. Video-based face recognition using probabilistic appearance manifolds. *IEEE Conference On Computer Vision and Pattern Recognition*, 1:313–320, 2003.

[LHYK05] K.C. Lee, J. Ho, M.H. Yang, and D. Kriegman. Visual tracking and recognition using probabilistic appearance manifolds. *Computer Vision and Image Understanding*, 99(3):303–331, 2005.

[Lib07] Liberty Human Right Organization. *http://www.liberty-human-rights.org.uk/news-and-events/1-press-releases/2007/britain-s-privacy.shtml*, 2007.

[LL07] N. Li and T. Li. t-closeness: Privacy beyond k-anonymity and l-diversity. In *IEEE International Conference on Data Engineering*, 2007.

[LMSR08] I. Laptev, M. Marszalek, C. Schmid, and B. Rozenfeld. Learning realistic human actions from movies. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8. IEEE, 2008.

[LSG97] A. Lee, K. Schlueter, and A. Girgensohn. Sensing activity in video images. In *CHI '97 extended abstracts on Human factors in computing systems: looking to the future*, pages 319–320, 1997.

[MKGV07] A. Machanavajjhala, D. Kifer, J. Gehrke, and M. Venkitasubramaniam. l-diversity: Privacy beyond k-anonymity. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 1(1):3, 2007.

[MKT04] A. Maxiaguine, S. Kunzli, and L. Thiele. Workload characterization model for tasks with variable execution demand. In *Proceedings of Design, Automation and Test in Europe Conference and Exhibition*, volume 2, pages 1040 – 1045 Vol.2, feb. 2004.

[MLS09] M. Marszalek, I. Laptev, and C. Schmid. Actions in context. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 2929–2936. IEEE, 2009.

[MO06] T. McGhee and L. Overley. GPS technology tracks employees. *The Denver Post, dec*, 2006.

[MPDMD05] I. Martínez-Ponte, X. Desurmont, J. Meessen, and J.F. Delaigle. c. In *International Workshop on Image Analysis for Multimedia Interactive Services*, 2005.

[MSS08] D. Mahapatra, M. Saini, and Y. Sun. Illumination invariant tracking in office environments using neurobiology-saliency based particle filter. In *IEEE International Conference on Multimedia and Expo*, pages 953 –956, 2008.

[MVW08] S. Moncrieff, S. Venkatesh, and G. West. Dynamic privacy assessment in a smart house environment using multimodal sensing. *ACM Transactions on Multimedia Computing, Communications, and Applications*, 5(2):1–29, 2008.

[NAC07] M. E. Nergiz, M. Atzori, and C. Clifton. Hiding the presence of individuals from shared databases. In *ACM International Conference on Management of Data*, pages 665–676, 2007.

[NAS08] M.E. Nergiz, M. Atzori, and Y. Saygin. Towards trajectory anonymization: a generalization-based approach. In *ACM GIS International Workshop on Security and Privacy in GIS and LBS*, pages 52–61. ACM, 2008.

[NIS10] NIST. Trec video retrieval evaluation (trecvid), 2001-2010. http://www-nlpir.nist.gov/projects/trecvid/.

[NSM05] E.M. Newton, L. Sweeney, and B. Malin. Preserving privacy by de-identifying face images. *Knowledge and Data Engineering*, 17(2):232 – 243, 2005.

[PCH09] J. K. Paruchuri, S. Cheung, and M. W. Hail. Video data hiding for managing privacy information in surveillance systems. *SPIE Newsroom*, 2009.

[Pen99] RM Pendyala. Measuring day-to-day variability in travel behavior using GPS data. *Final Report, FHWA, Washington, DC, URL: http://www. fhwa. dot. gov-/ohim/gps*, 1999.

[PET11] PETS. Performance evaluation of tracking and surveillance, 2000-2011. http://www.cvg.cs.rdg.ac.uk/slides/pets.html.

[PF11] C. Piciarelli and G.L. Foresti. Surveillance-oriented event detection in video streams. *IEEE Intelligent Systems*, pages 32–41, 2011.

[PGM10] X. Pan, Y. Guo, and A. Men. Traffic surveillance system for vehicle flow detection. In *International Conference on Computer Modeling and Simulation*, pages 314–318, 2010.

[PKV07] C. Patrikakis, P. Karamolegkos, and A. Voulodimos. Security and privacy in pervasive computing. *Pervasive Computing, IEEE*, 6(4):73 –75, 2007.

[Qur09] F. Z. Qureshi. Object-video streams for preserving privacy in video surveillance. In *International Conference on Advanced Video and Signal Based Surveillance*, pages 442–447, 2009.

[R.08] Steven R. Privacy is dead, get over it. In *The Last Hope conference*, 2008.

[Rat10] TD Raty. Survey on contemporary remote surveillance systems for public safety. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, 40(5):493–515, 2010.

[SAM+10] M. Saini, P. Atrey, S Mehrotra, S Emmanuel, and M. Kankanhalli. Privacy modeling in video data publication. In *IEEE International Conference on Multimedia and Expo*, pages 60–65, 2010.

[SEY05] B. Song, C. Ernemann, and R. Yahyapour. Parallel computer workload modeling with markov chains. In *Job Scheduling Strategies for Parallel Processing*, pages 9–13. Springer, 2005.

[SG99] C. Stauffer and W.E.L. Grimson. Adaptive background mixture models for real-time tracking. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 2, page 252 Vol. 2, 1999.

[SMM+09] J. Schiff, M. Meingast, D.K. Mulligan, S. Sastry, and K. Goldberg. Respectful cameras: Detecting visual markers in real-time to address privacy concerns. *Protecting Privacy in Video Surveillance*, pages 65–89, 2009.

[SOK06] A. F. Smeaton, P. Over, and W. Kraaij. Evaluation campaigns and trecvid. In *ACM International Workshop on Multimedia Information Retrieval*, pages 321–330, 2006.

[SPH+05] A. Senior, S. Pankanti, A. Hampapur, L. Brown, Ying-Li Tian, A. Ekin, J. Connell, Chiao Fe Shu, and M. Lu. Enabling video privacy through computer vision. *Security Privacy, IEEE*, 3(3):50 – 57, 2005.

[SQGP06] K. Smith, P. Quelhas, and D. Gatica-Perez. Detecting abandoned luggage items in a public space. In *International Workshop on Performance Evaluation in Tracking and Surveillance*, pages 75–82, 2006.

[STT06] H. Septian, Ji Tao, and Yap-Peng Tan. People counting by video segmentation and tracking. In *International Conference on Control, Automation, Robotics and Vision*, pages 1–4, 2006.

[Swe02] L. Sweeney. k-anonymity: A model for protecting privacy. *International Journal on Uncertainty Fuzziness and Knowledgebased Systems*, 10(5):557–570, 2002.

[SWH+06] T. Spindler, C. Wartmann, L. Hovestadt, D. Roth, L. Van Gool, and A. Steffen. Privacy in video surveilled areas. In *The ACM International Conference on Privacy, Security and Trust*, pages 1–10, 2006.

[TH01] S. Tansuriyavong and S. Hanaki. Privacy protection by concealing persons in circumstantial video image. In *ACM Workshop on Perceptive user interfaces*, pages 1–4, 2001.

[TIR94] J. C. Tang, E. A. Isaacs, and M. Rua. Supporting distributed groups with a montage of lightweight interactions. In *ACM conference on Computer supported cooperative work*, pages 23–34, 1994.

[TLB+06] B. Thuraisingham, G. Lavee, E. Bertino, J. Fan, and L. Khan. Access control, confidentiality and privacy for video surveillance databases. In *ACM Symposium on Access Control Models and Technologies*, pages 1–10, 2006.

[TM08] M. Terrovitis and N. Mamoulis. Privacy preservation in the publication of trajectories. In *International Conference on Mobile Data Management*, pages 65–72. IEEE, 2008.

[VHLP08] F. Van Harmelen, V. Lifschitz, and B. Porter. *Handbook of knowledge representation.* Elsevier Science Ltd, 2008.

[WBSS04] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: From error visibility to structural similarity. *IEEE Transaction on Image Processing*, 13(4):600–612, 2004.

[WDMV04] J. Wickramasuriya, M. Datt, S. Mehrotra, and N. Venkatasubramanian. Privacy protecting data collection in media spaces. In *International Conference on Multimedia*, pages 48–55, 2004.

[WJ07] U. Westermann and R. Jain. Toward a common event model for multimedia applications. *IEEE MultiMedia*, 14(1):19–29, 2007.

[WR11] T. Winkler and B. Rinner. Securing embedded smart cameras with trusted computing. *EURASIP Journal on Wireless Communications and Networking*, page 8, 2011.

[ZCC05] W. Zhang, S. Cheung, and M. Chen. Hiding privacy information in video surveillance system. In *IEEE International Conference on Image Processing.*, volume 3, pages II – 868–71, 2005.

[ZCZM10] F.W. Zhu, S. Carpenter, W. Zhu, and M. Mutka. A Game Theoretic Approach to Optimize Identity Exposure in Pervasive Computing Environments. *International Journal of Information Security and Privacy*, 4(4):1–20, 2010.

[ZS98] Q.A. Zhao and J.T. Stasko. Evaluating image filtering based techniques in media space applications. In *ACM conference on Computer supported cooperative work*, pages 11–18, 1998.