Hindawi Publishing Corporation
Advances in Multimedia
Volume 2012, Article ID 639649, 14 pages
doi:10.1155/2012/639649

Research Article

Adaptive Transformation for Robust Privacy Protection in Video Surveillance

Mukesh Saini,[1] Pradeep K. Atrey,[2] Sharad Mehrotra,[3] and Mohan Kankanhalli[1]

[1] *School of Computing, National University of Singapore, Singapore 117417*
[2] *Department of Applied Computer Science, The University of Winnipeg, MB, Canada R3T 5V9*
[3] *Information and Computer Science Department, University of California, Irvine, CA 92697-3425, USA*

Correspondence should be addressed to Mukesh Saini, mksaini@comp.nus.edu.sg

Received 30 November 2011; Revised 6 February 2012; Accepted 6 February 2012

Academic Editor: Martin Reisslein

Privacy is a big concern in current video surveillance systems. Due to privacy issues, many strategic places remain unmonitored leading to security threats. The main problem with existing privacy protection methods is that they assume availability of accurate region of interest (RoI) detectors that can detect and hide the privacy sensitive regions such as faces. However, the current detectors are not fully reliable, leading to breaches in privacy protection. In this paper, we propose a privacy protection method that adopts adaptive data transformation involving the use of selective obfuscation and global operations to provide robust privacy even with unreliable detectors. Further, there are many implicit privacy leakage channels that have not been considered by researchers for privacy protection. We block both implicit and explicit channels of privacy leakage. Experimental results show that the proposed method incurs 38% less distortion of the information needed for surveillance in comparison to earlier methods of global transformation; while still providing near-zero privacy loss.

1. Introduction

In order to perform privacy-preserving CCTV monitoring, video data should be transformed in such a way that the information leaking the identity is hidden, but the intended surveillance tasks can be accomplished. The traditional approach of data transformation has been to detect the regions of interest (RoI) in the images (e.g., human faces) and selectively obfuscate them. This approach is an unreliable solution as the RoI detectors may sometimes fail. For example, even if a face detector is able to correctly detect the face in 99 (out of 100) frames, the undetected faces in the remaining frame will reveal the identity of the person in the video and result in his/her privacy loss.

In other set of works, global operations have been used for data transformation in which the whole video frame is transformed with same intensity, that is, same amount of blurring or quantization [1]. This approach is more appropriate in the context of data publication, where the published surveillance video is used by researchers for testing their algorithms. In contrast to the data publication scenario, CCTV monitoring scenario has different requirements. In the case of CCTV monitoring, a human operator is required to watch the surveillance video feeds; although automated techniques may run in the background as shown in Figure 1. The automatic analysis can be performed using the original data, which is not accessible for viewing, unlike data publication. The original data may be encrypted and stored in a database which can be retrieved later in the event of emergency situations. The RoI information obtained using the detectors, along with the transformed data, can be presented to the human operators. Further, the RoI information can be used to adapt data transformation. We take this opportunity to explore an adaptive data transformation approach to combine the benefits of both selective obfuscation and global operations.

In this paper, to overcome the nonreliability of the RoI detectors, we examine the suitability of an adaptive approach of data transformation in order to provide near-zero privacy loss in a CCTV monitoring scenario. In the proposed privacy protection framework, data transformation is performed in two stages. In the first stage, automatic detectors (mainly

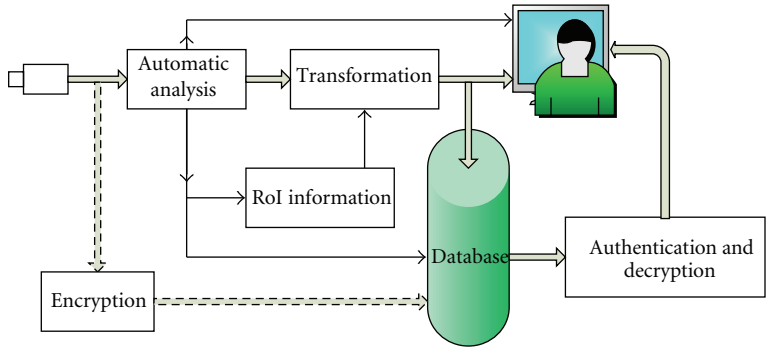


Figure 1: The automatic algorithms run on the original data. A transformed version can be showed to the CCTV operators.

blob and face detectors) are applied on the data for detection of evidences. The results from these detectors are used to adapt the global operation. The adaption is done in two dimensions: spatial (by using a space variant operation) and temporal (by providing a failure time window to the detectors). For privacy loss assessment, we adopt the model proposed in [2], as it considers both implicit and explicit identity leakage channels.

The main contributions of the paper are the following:

(i) an adaptive data transformation approach that uses space variant operations is proposed, which provides a near-zero privacy loss with minimal visual distortion;

(ii) the proposed method provides robust privacy preservation even with low-accuracy detectors.

Rest of the paper is organized as follows. We compare proposed work with previous works in Section 2. In Section 3, we describe proposed privacy protection method. Experimental results are analyzed in Section 4, and paper is concluded in Section 5.

2. Related Work

Most researchers [3–10] have used selective obfuscation to preserve privacy in 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 detectors, privacy cannot be guaranteed.

In Table 1, we present a comparison of the proposed work with other works in the following aspects: whether implicit identity leakage channels (e.g., location, time, and activity information) have been used for assessing privacy loss; whether a tradeoff between privacy loss and visual distortion of the whole frame due to data transformation has been examined, and which of the approaches (selective obfuscation or global operations) has been adopted. As shown in Table 1, 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, the proposed privacy preserving method presents a tradeoff between utility and privacy in a given CCTV monitoring scenario. Finally, the proposed method examines an adaptive approach for data transformation. We use face and blob detectors to detect the regions in the image that need to be obfuscated. However, the inaccuracies of these detectors is overcome by adapting operations spatially as well as temporally in the video.

We differentiate the contributions in the paper from our past work [2] as follows. While in [2], we introduced the notion of implicit identity leakage channels and provided a computational model for identity leakage and privacy loss, in this paper we examine the appropriateness of data transformation operations in order to block these identity leakage channels. One approach could be to globally transform

Table 1: A comparison of the proposed work with the existing works.

The work	Implicit identity leakage channels used?	Utility/privacy tradeoff?	Approach adopted
Boyle et al. [1]	No	No	GO
Senior et al. [3]	No	No	SO
Moncrieff et al. [11]	No	No	SO
Fidaleo et al. [4]	No	No	SO
Wickramasuriya et al. [5]	No	No	SO
Koshimizu et al. [6]	No	No	SO
Spindler et al. [12]	No	No	SO
Thuraisingham et al. [7]	No	No	SO
Carrillo et al. [8]	No	No	SO
Paruchuri et al. [9]	No	No	SO
Qureshi [10]	No	No	SO
Saini et al. [2]	Yes	No	No transformation
Proposed work	Yes	Yes	SO and GO

SO: selective obfuscation; GO: global operations.

the data to provide a tradeoff between the privacy loss and utility loss. However, global data transformation operations are not appropriate in a CCTV monitoring scenario since the global operations introduce large amount of visual distortions in the video. Therefore, in this paper we propose an adaptive data transformation approach that combines benefits of selective obfuscation and global operations to provide robust privacy with minimal distortion. Further, we provide a tradeoff between the visual distortions due to data transformation and the privacy loss of the people present in the video.

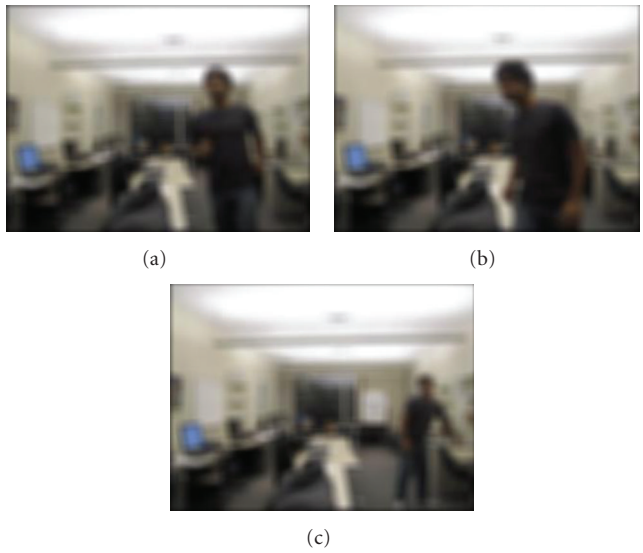
3. Privacy Protection Method

In the previous works, it is identified that the identity leakage and privacy loss occur due to presence of the evidences information such as *who, what, when,* and *where*. An evidence can be learned from multiple sources. For example, the *where* evidence can be detected using text legends, familiar structures, symbols (company logos), and so forth. In order to robustly block the identity leakage, we need to remove/modify all the sources of evidence detection. In a surveillance scenario, relatively static regions of the camera view are known as the background, for example, rigid structures, fixed objects, doors, and exits. Anything that is not background is considered foreground, which generally corresponds to the humans walking in the camera view [13]. We observed that some of the sources of identity leakage are found in the background, while others are part of the foreground/object itself. Particularly, most of the sources of *where* and *when* evidences are embedded in the background; while the *who* and *what* evidences are usually found in the foreground. Further, we observe that the surveillance cameras are generally fixed, resulting in static background [14]. Since the background is mostly static, the sources which are part of background can be accurately detected manually and transformed. The foreground parts need to be automatically removed as they may appear at varying places in different frames.

Most sources of the evidences can be associated to a region in the image called evidence regions. For example, a rectangle encompassing a company logo, that can provide the company and its location information, is one evidence region for *where* evidence. Our aim is to transform the image such that all the evidence regions are obscured enough to block the identity information. However, the problem is that we may not be able to detect these regions accurately due to the limitations of the automatic techniques [15, 16]. In the proposed method, these inaccuracies are taken care of by using spatially and temporally adaptive data transformation. The quality of the transformed data is measured in terms of perceptual distortion D, which is computed as

$$D = 1 - \text{SSIM}, \tag{1}$$

where SSIM is the structural similarity index [17]. We use SSIM value over PSNR because this measure is more consistent with human eye perception [18, 19]. For the sake of completeness, we first provide a brief overview of selective obfuscation and global operations and then describe the proposed method.



(a) (b) (c)

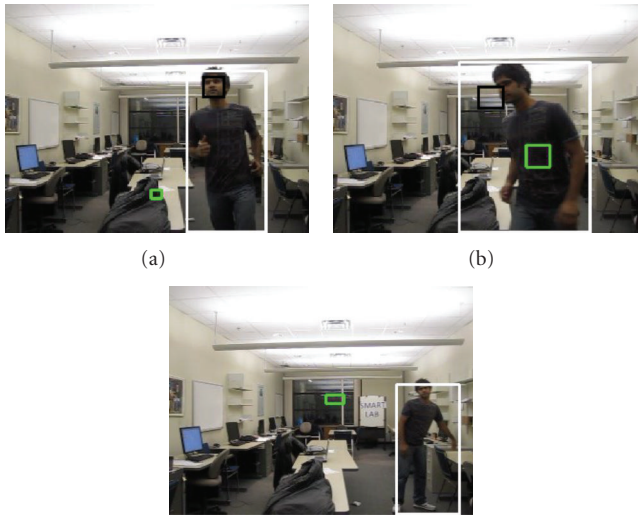
Figure 2: The images are blurred to hide the identity information.

3.1. Existing Approaches

3.1.1. Selective Obfuscation. In these methods, the evidence revealing image regions are selected using computer vision techniques and subsequently obfuscated. For example, Figure 3 shows the results of face detection for hiding the facial information. In the first image the face is detected properly, which helped in accurately removing the facial information. However, in the second image the face regions are incorrectly detected, while in the third image they are not detected at all. Note that if the face is left undetected and seen in even one frame, the identity is revealed. Hence, selective obfuscation methods do not provide robust privacy preservation.

3.1.2. Global Operations. To overcome the problem of unreliable vision algorithms, we can perform generalization on the whole image. For instance, we can coarsely quantize the image, scramble the color space, or blur the image. The problem with these methods is that they are generally too pessimistic; we need to determine the worst case transformation parameters (e.g., degree of blurring or quantization) and blur all the images to that amount, irrespective of the content of the image. This is in contrast to the fact that when the person is far from the camera, even little blurring might be sufficient. Figure 2 shows the result of this approach where the images are blurred to hide faces. From this figure, we observe that the image background gets distorted even when the object occupies a small portion of the image. The background information might be important for a surveillance person in order to understand the situation.

3.2. Proposed Adaptive Approach. We propose an adaptive method that uses global transformation according to the results of selective obfuscation. In this method, we first use



(a) (b) (c)

Figure 3: The results of the blob detection (white rectangle) and face detection (black rectangle).

face and blob detectors to approximate the location of the persons in the image and then use *space variant operation* to hide the identity. Figure 3 shows the results of blob detection and face detection on the same data. We can observe that a blob detector is generally more robust than the a face detector in detecting the presence of a person; although the boundaries may not be very accurate, we can still get a good approximation of the centroid of the region occupied by the person. This centroid information is used to perform a space variant operation. In the space variant operation, the operation parameters vary with the space according to a profile. Let r_i be the image region with evidence information and c_i the most probable (for face presence) point of the region, then the quality $\mathcal{Q}$ for a pixel p in region r_i is calculated as follows:

$$\mathcal{Q}(p) = \mathcal{Q}_0 f(\Delta(c_i, r_i)), \tag{2}$$

where f is the profile function, $\mathcal{Q}_0$ is the operation parameter for the centroid, and Δ is a distance function. In a ramp profile, for example, the transformation intensity decreases linearly with the distance from the centroid. This mechanism has the advantage that even if the approximate location of a person is determined, the evidence regions can be obscured with high probability without globally transforming the image.

The space variant operations are useful when the detectors are unable to detect the evidence regions correctly, but only provide an approximation of the region which can cause evidence detection. However, sometimes the detectors completely fail to detect persons in the video. We analyzed the failure pattern of the blob detector over a number of videos and made the following observations:

(i) when the person enters the camera view, the background model-based blob detector detects the person reliably;

(ii) the detector may fail to detect a person due to noise or lighting changes, and so forth;

(iii) the maximum number of contiguous frames in which the blob detector fails is limited.

One such failure pattern is shown in Figure 4 in which the person enters the camera view in 71th frame and leaves in 910th frame. The detector fails for the following frames: 220 to 230, 460 to 495, 710 to 740, and 810 to 830. To model this failure pattern, we define a failure window ω. If the number of contiguous frames in which the blob is not detected is less than ω, we assume that the person is present in the video, but the blob detector has failed to detect that person. In this situation we adopt the pessimistic approach and globally transform the whole image. If no blob is detected for more than ω contiguous frames, we conclude that the person has left the camera view and there is no need for a global transformation.

Note that our aim is to reduce the privacy loss when the data is presented to the surveillance operator for viewing. The automatic algorithms can still work on the original data, but human beings can only see the transformed data. Nonetheless, in emergency situations, a person with authority can access the original data. Figure 1 shows how the proposed method can be deployed in a surveillance system.

3.2.1. Description of Algorithm. The evidence regions can be divided into two groups:

(i) *static evidence regions:* these are the regions of the background that provide us evidence which can lead to identity. Let $R^s = \{r_1^s, r_2^s \ldots\}$ be the set of background evidence regions, which include any text legends, landmark or famous buildings, name plates, addresses, symbols and logos, and so forth;

(ii) *dynamic evidence regions:* these are the foreground regions that provide *who* and *what* evidence. Let $R^d = \{r_1^d, r_2^d, \ldots\}$ be the set of image regions detected as foreground using blob and face detectors. Each region is defined by a centroid, width, and height. Dynamic evidence regions may vary with time; therefore, these are calculated on-the-fly for the current frame.

The proposed method is described in Algorithm 1. The algorithm takes a video V and set of static evidence regions R^s as input and returns the transformed video V'. The important steps of the algorithm are explained below.

Statement 3. The function *DSR*() estimates the dynamic evidence regions using blob and face detector. To detect the evidence regions, we tune the thresholds of the detectors to minimize the number of false negatives. In the experiments we show that we are able to obtain very low number of false negatives.

Statements 4 to 13. If no foreground is detected, there can be two cases: (1) there is no foreground region in the image; (2) the detector failed to detect the foreground. Whether current

Input: Original Video: $V = \{f_1, f_2 \ldots\}$
and set of static evidence regions: $R^s = \{r_1^s, r_2^s \ldots\}$;
Output: Transformed Video: $V' = \{f_{1'}, f_{2'} \ldots\}$
Description:

```
1: for all f_i ∈ V do
2:   // Detect dynamic evidence regions
3:   R^d = DSR(f_i);
4:   if R^d == EMPTY then
5:     if R^d == EMPTY for previous ω frames then
6:       f_i^t = COPY(f_i);
7:       for all r_j^s ∈ R^s do
8:         f_i^t = ST(f_i^t, r_j^s);
9:       end for
10:    else
11:      //Do the global transformation
12:      f_i^t = GT(f_i);
13:    end if
14:  else
15:    //Foreground detected
16:    //Transform static regions
17:    f_i^t = COPY(f_i)
18:    for all r_j^s ∈ R^s do
19:      f_i^t = ST(f_i^t, r_j^s);
20:      MARK(f_i^t, r_j^s) = TRUE;
21:    end for
22:    //Transform dynamic regions
23:    for all frame r_k^d ∈ R^d do
24:      //Calculate parameters for space variant operation
25:      PRM = PE(r_k^d)
26:      if !MARK(f_i^t, r_k^d) then
27:        f_i^t = DT(f_i^t, r_k^d, PRM);
28:      end if
29:    end for
30:    //Data transformation over
31:  end if
32:  //Copy frame to output frame sequence
33:  f_i' = COPY(f_i^t);
34: end for
35: return Transformed frame sequence f_i' as Video V';
```

Algorithm 1: Adaptive data transformation.

frame belongs to the first case or the latter case is determined by examining the $DSR()$ output for the previous ω frames. If we do not detect any foreground regions in the previous ω frames, then it is very likely that there is no person in the image; hence, we only transform the static regions. On the other hand, if some foreground is detected within last ω frames, there are more chances of existence of a person in the image. In this case, we take the pessimistic approach and globally transform the whole image. In the function $ST()$, the static evidence regions of the images are obscured using a suitable transformation operation. An evaluation of three operations, namely, blurring, pixelization, and quantization is provided in the experimental results. Similarly, $GT()$ transforms the entire image globally.

Statements 17 to 21. When the foreground is detected, we selectively transform static and dynamic evidence regions. In this case, we first do the static transformation and then pass the image for transformation of dynamic evidence regions.

Statements 23 to 29. In these steps we transform the dynamic regions of the image. Only those regions are selected for the transformation that are not obscured during the static transformation. The dynamic transformation is done in two steps. In the first step, parameters of the dynamic region r_i^d (centroid, height, and width) are used to estimate ($PE()$) a probable area (PRM) (circular in our case), where the evidence could exist. The details of the parameter estimation are discussed in the experiments section. This area is then space-variantly transformed according to the probability of finding evidence information, that is, the subareas where the probability of finding evidence is less, they are transformed with less degree and vice versa. This space variant transformation operation (e.g., blurring, pixelization, etc.) is

performed in function $DT()$; an implementation of which will be discussed with the experiments.

Space variant operations incorporate operating context in data transformation. For example, if the detectors being used are less accurate, a bigger area can be selected for dynamic transformation. Since the degree of transformation decreases with the distance from the center, we do not compromise much in the quality. By analyzing the frames over a temporal window and selecting a proper transformation function, we are able to accommodate temporary failures of the detector.

4. Experimental Results

We performed five experiments to demonstrate the efficacy of the proposed privacy protection method. In the first experiment we highlight the effect of nonfacial information on privacy loss. We also provide an evaluation of blurring, pixelization, and quantization transformation operations that are required to remove the static evidence regions, which provide *when* and *where* evidences. In the second experiment, we show the improved visual quality obtained using the proposed method for a near-zero privacy loss. It is shown that the proposed method that adaptively uses selective obfuscation and global operations is more reliable than the selective obfuscation alone, and it achieves better quality than the global transformation alone. It is also demonstrated how the spatial and temporal adaption can be used to overcome the inaccuracies of the detectors. In Experiment 3, we analyze how privacy loss and visual distortion are affected by varying ω. An attempt to improve the proposed method is made in Experiment 4 to overcome a special failure pattern that might occur in a multiperson scenario. Finally, we validate our conclusions with an experiment on 24 hours of real surveillance data in Experiment 5.

4.1. Data Set. Five video clips have been considered in our experiments. The description of the video clips is as follows:

(i) video 1: this video was recorded in a research lab. It shows name of the lab and two people doing various activities. The original video was shot for over one hour consisting of 200 key frames;

(ii) video 2: the video is recorded at the entrance of a department building. It has multiple *where* evidences in the form of text and logo. The video is of 45 minutes length, and it consists of 483 frames;

(iii) video 3: this is again a video recorded at a research lab where two people are doing some activities. It consists of 1095 frames;

(iv) video 4: this video was shot at the wash basin in a canteen. Two people are seen in the video at a time and it consists of 1520 frames;

(v) video 5: the video consists of 1200 frames which are taken from PETS data sets [20].

Figure 5 shows the background images for the five video clips used in this experiment. From the figure it can be

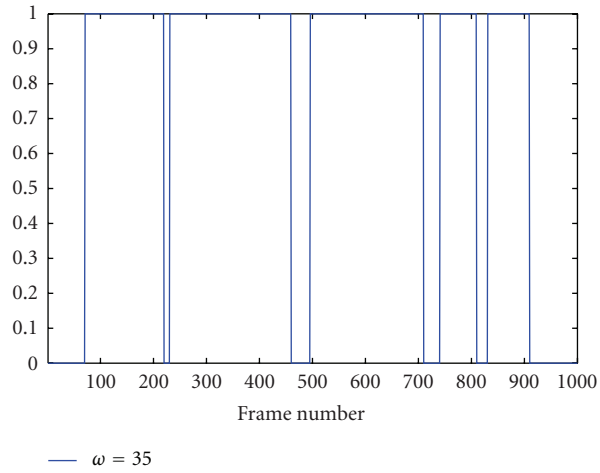


Figure 4: A failure pattern of blob detector, y-axis; 1 mean Detected correctly, 0 mean Failed.

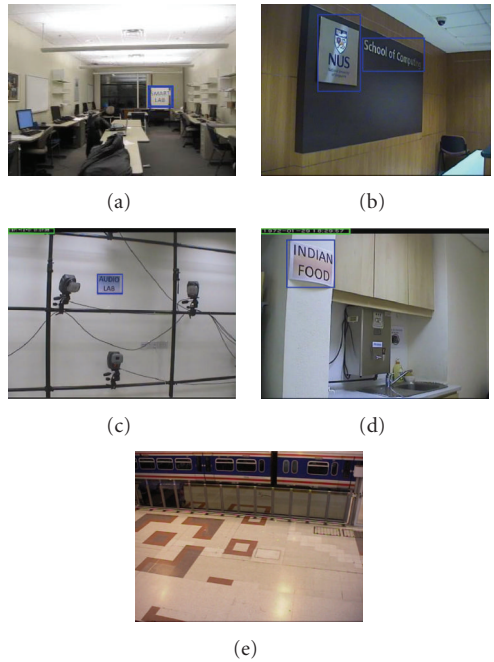


Figure 5: The background pictures for five videos. Green box shows *when* evidence and blue rectangle shows *where* evidence.

derived that: in video 1 and video 2, we can detect *what* and *where* evidences; in video 3 and video 4, we can detect *what*, *where*, and *when* evidences; but in video 5, we can only detect *what* evidence. To validate the conclusions in real scenarios, we also use 24 hours of real surveillance footage consisting of 28216 frames (video 6). The video frames are not shown due to privacy concerns. We can detect *what* evidence in the real video.

Experiment 1 (*implicit identity leakage and static regions*). A large amount of work assumes that privacy loss only occurs due to the presence of the facial information in the

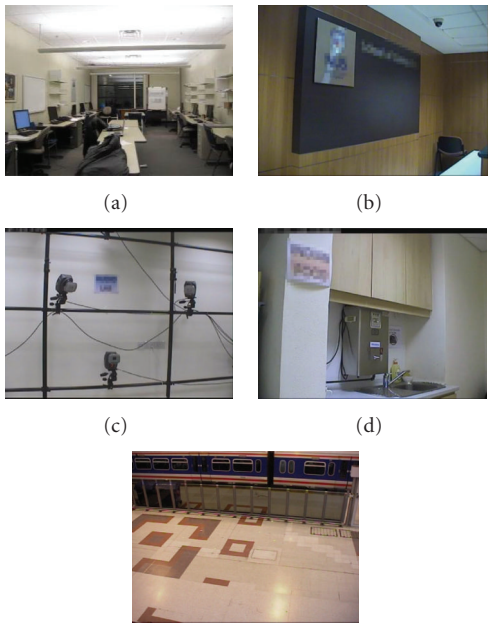


Figure 6: The static evidence regions after pixelization (Video 1–3, Video 2–8, Video 3–6, Video 4–9). Video 5 does not have any evidence region in background.

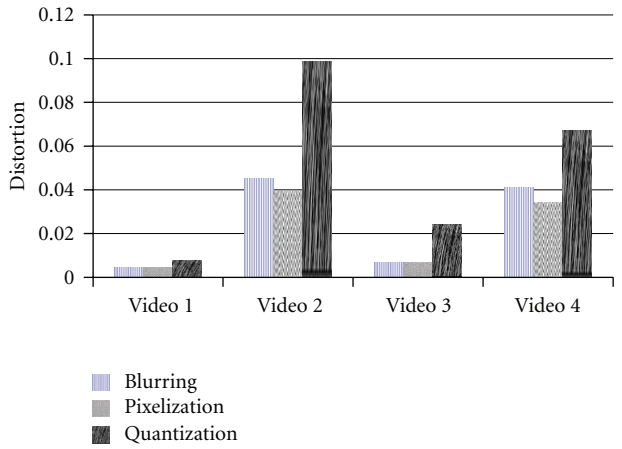


Figure 7: Different transformation operations to obscure the static evidence regions and corresponding distortion measures. Video 5 does not have any static evidence regions.

image. In this experiment we mainly highlight the limitations of the earlier privacy protection methods and show how the evidences found in the image can cause privacy loss even without facial information. To highlight the effect of the implicit channels alone, in this experiment we assume that the face is already removed from the videos and then calculate the privacy loss based on the model described in [2]. The associated cluster sizes depend on the scenario in which the video is recorded; however, for experimental purposes, we take the following values of clusters: $C_{\text{what}} = 10000$, $C_{\text{what,when}} = 3000$, $C_{\text{what,where}} = 20$, and $C_{\text{what,when,where}} = 5$.

Table 2: The privacy loss calculation for different video clips.

Video	Evidences	Identity leakage	Privacy loss
Video 1	*What; where*	$I_{\text{what,where}} = 2/20$	0.119
Video 2	*What; where*	$I_{\text{what,where}} = 1/20$	0.018
Video 3	*What; where; when*	$I_{\text{what,where,when}} = 2/5$	0.880
Video 4	*What; where; when*	$I_{\text{what,where,when}} = 1/5$	0.119
Video 5	*what*	$I_{\text{what}} = 6/10000$	0.002

In Table 2 we present the privacy loss that might occur from these video clips even when the face is not present.

It can be observed that if the adversary has the prior knowledge of the clusters, the nonfacial information can also cause significant privacy loss, and, therefore, we need to remove these evidences from the videos to minimize the privacy loss. We explore three operations to transform the static evidence regions: blurring, quantization, and pixelization. We perform these operations to the degree that the evidence is not detectable and compare the perceptual distortion they cause. The static evidence regions of the videos are shown as green and blue boxes in Figure 5; note that video 5 does not have any static evidence region. In Figure 7 we notice that pixelization performs equivalent or better than blurring and quantization operations. On average over all the videos, pixelization incurs 8% less distortion than blurring and 55% less distortion than quantization. The resulting pixelized images are shown in Figure 6. In the remaining experiments, we will use pixelization to obscure static evidence regions. This experiment only removes evidences from static background, we still need to consider *who* and *what* evidences which can be learned from the dynamic foreground.

Experiment 2 (*space variant operation*). In this experiment we examine the use of the techniques to remove evidences that are detected from the foreground. As mentioned before, the most common evidences that are found in the foreground of video frames are *what* (activities in our case) and *who* (face in our case). The identity leakage through *what* alone is negligible, hence we put more focus on the facial information removal. Also note that removing *what* evidence can severely affect the intended surveillance objective.

One extreme solution to overcome the nonreliability of the detectors is to globally transform the image. For example, we can blur the whole image irrespective of the location of the face. To evaluate this technique, we applied the operations on the whole video to the extent, where the face became unrecognizable in all the frames. Figure 8 shows the results of blurring and pixelization on the five videos. We observe that except for Video 5, blurring performs better than pixelization. This is probably because the distortion in the case of pixelization increases more rapidly compared to blurring as the faces are captured more close to the camera. The dips in the plots show the regions of high activity with multiple people. However, it cannot be avoided as the probability of privacy loss is also higher in those frames.

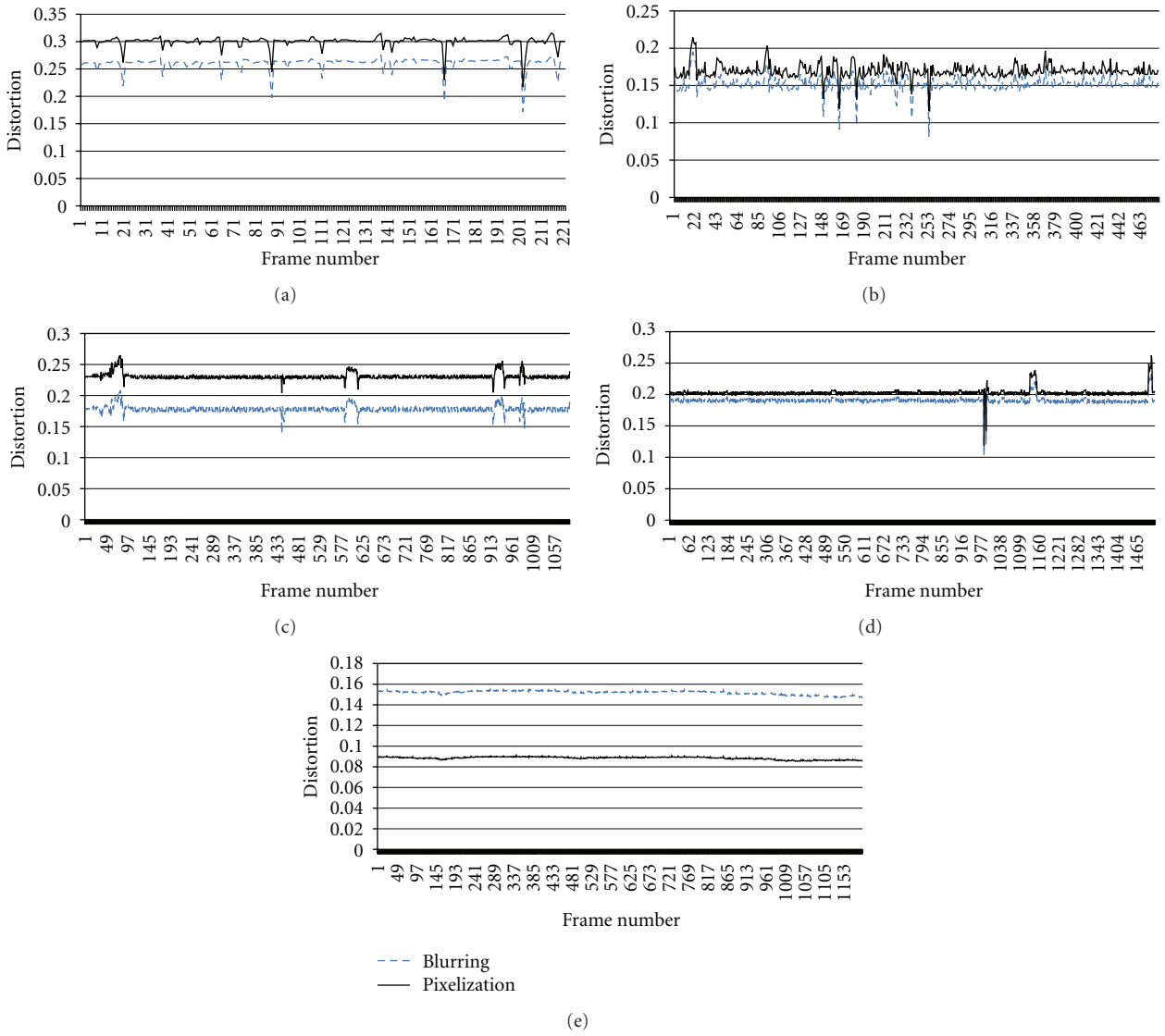


Figure 8: The distortion measures for blurring and pixelization using global transformation.

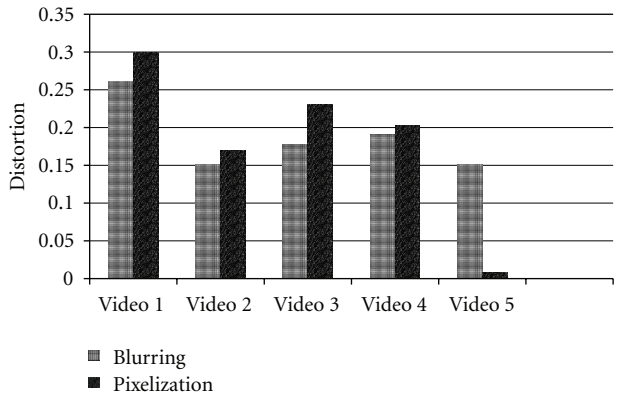


Figure 9: Comparison of distortion due to blurring and pixelization for global transformation. If P is the degree of pixelization, and B is degree of blurring, we got following values: video 1: $B = 6$ and $P = 4$, video 2: $B = 13$ and $P = 8$, video 3: $B = 12$ and $P = 8$, video 4: $B = 12$ and $P = 8$, and video 2: $B = 6$ and $P = 3$.

The overall comparison of average distortion values for all five videos is shown in Figure 9.

The foreground regions only occupy a small part of the image, hence uniformly transforming the whole image is a very pessimistic approach. To overcome the non reliability of the face detectors, we propose to use more robust foreground detectors (e.g., blob detector) which can be made very reliable by reducing the threshold values, although at the cost of increased false positives. We conducted experiments with GMM-based adaptive background modeling to detect the blobs [21]. The results of the blob detection are shown in Figure 10. By reducing the threshold values we are able to detect foreground in most of the images.

However, transforming only the blob regions has two problems: (1) in some frames the blob may include the body of the person but still miss the face, for example, in Figure 10(a) and (2) the face only occupies small region of the blob, hence transforming whole blob region may be too

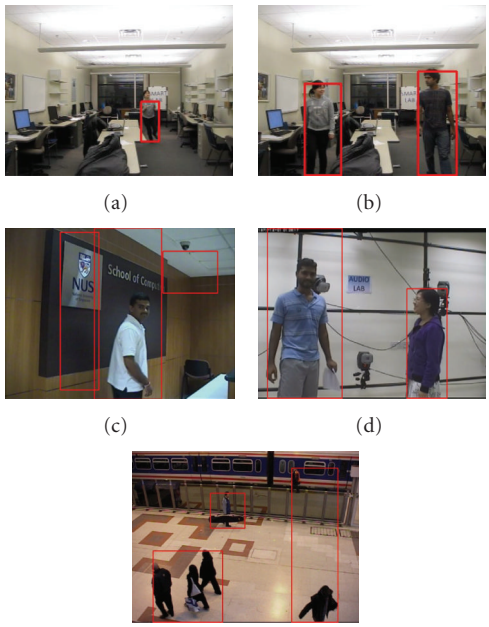


Figure 10: Results of blob detection on five videos. The blob detection works in the image where face detector fails.

pessimistic. From blob detection results and global transformation of the images, we make the following observations. (1) we need to apply more blurring/pixelization to obscure the frontal faces; however, the degree of blurring/pixelization could be less when the person is not directly looking at the camera. (2) the frontal face is generally found at 75% height of the blob. (3) the missed faces are within 125% of the height of the blob. These observations inspire us to use a space variant transformation, where the degree of transformation varies with the distance from the center of the estimated facial region.

In the implemented space variant transformation, a circular evidence region is estimated based on the inaccuracy of the detectors, and then different regions of the circle are transformed by different degrees of the transformation considering the distance from the center c_i (according to (2)). Based on the observations mentioned in the previous paragraph, the center (c_i) is determined as $(B_x + (3/2) * B_h, B_y + (1/2) * B_w)$, where (B_x, B_y) are the coordinates of the bottom-left corner of the blob with respect to the bottom-left corner of the image, and (B_w, B_h) are width and height of the blob. The radius of the circle is approximated as $(\max((1 + \mu) * B_h, (1 + 2 * \mu) * B_w))$, where μ is the fractional error margin which is 0.25 for the blob detector in our case. A less accurate blob detector would need higher value of margin μ. The circular region obtained above is divided into four concentric circles. The value of $\mathcal{Q}_0$, that is, the transformation parameter for the innermost circle is chosen according to the results of the global transformation (Figure 9) as follows: $\mathcal{Q}_0$ for blurring- 6, 13, 12, 12, and 6; for pixelization- 4, 8, 8, 8, and 3, respectively for five videos. The profile function f is chosen to be piecewise linear, and the function Δ is based on the Euclidean distance. With each

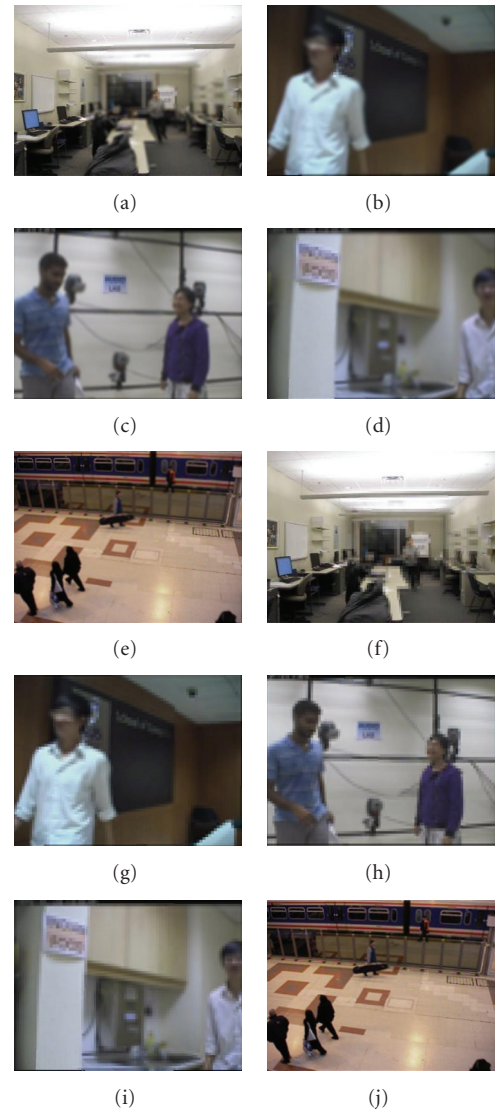


Figure 11: Key images from the transformed video using the proposed method. Row 1 shows the outputs of blurring, whereas row 2 shows the pixelization outputs.

outer circle, the blurring parameter is reduced by 5% and rounded to the nearest integer. The face detector output is also used to provide additional robustness. The implemented face detector provides a square facial region. In this case, the center is calculated as (Side/2, Side/2), and radius is taken as $\max((1 + 2 * \mu) * \text{Side})$ to account for inaccuracy.

Again, these numbers may depend on the context of the surveillance and accuracy of the detectors. In the current experimental settings, these parameters are obtained to give near-zero privacy loss (no face recognition possible from the transformed data) for given videos and blob detector.

Now we evaluate the proposed method from the perspective of visual distortion. For evaluation of the proposed method, we implement Algorithm 1 (described in Section 3.2). Figure 11 shows the resulting output images. The resulting values of the distortion for video clips (of Figure 11) are shown in Figure 12. The variation in the

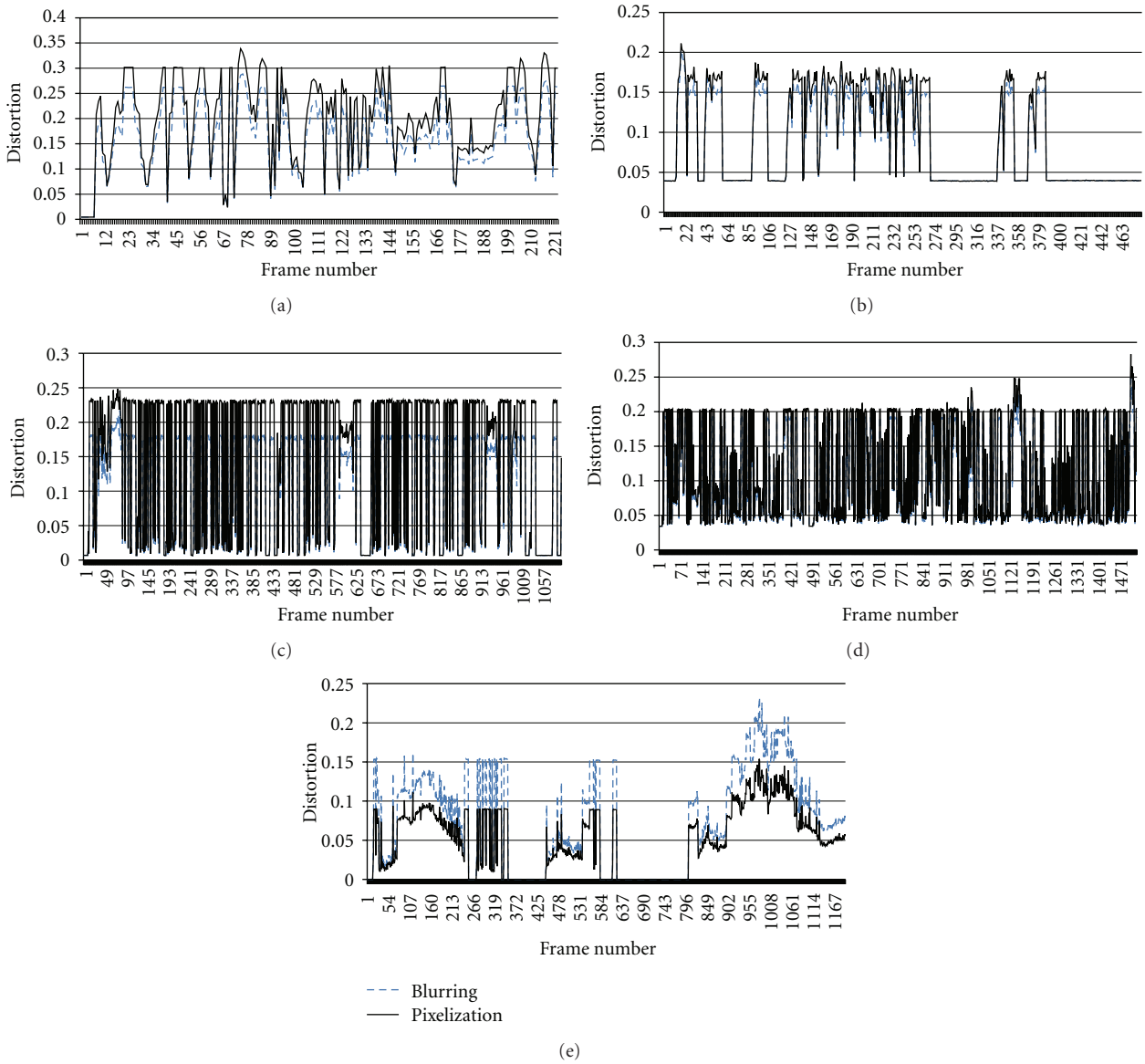


Figure 12: The distortion measures for blurring and pixelization using the proposed method.

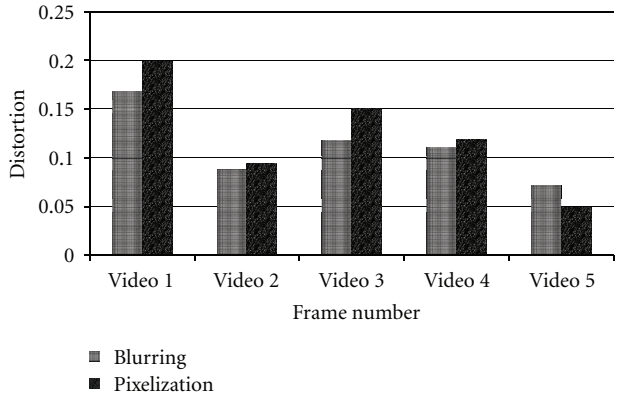


Figure 13: A comparison of distortion measure for pixelization and blurring for all video clips using the proposed method.

distortion is much more in comparison to global transformation. This is because when no blobs are detected, only static regions of the video are transformed; resulting in low distortion. On the other hand, sometimes blobs of large size are detected (probably due to increased false positive rate), which cause whole image to be transformed. It can be observed that even when the whole image is transformed, the distortion value using space variant method is less. This is due to the fact that in space variant operations less probable evidence regions are transformed mildly.It can be observed from Figure 13 that blurring provides a more effective solution for transforming the foreground regions. In Figure 14, a comparison of the proposed method with the global transformation is provided. The results show that we get 37% less distortion with proposed method in comparison to global method, still providing robust privacy protection.

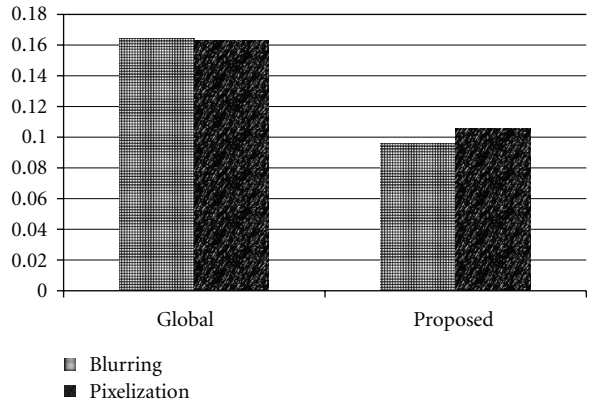


Figure 14: A comparison of global method and proposed method.

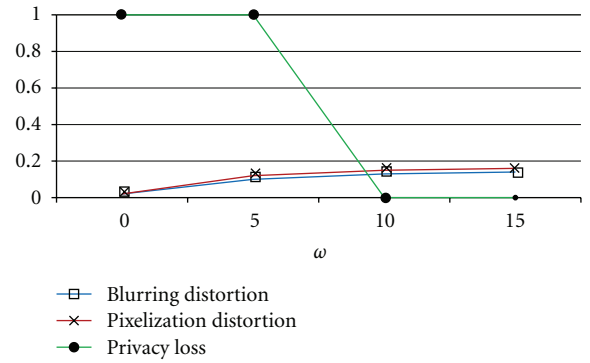


Figure 15: The effect of ω on privacy loss and visual distortion.

For the given video clips, we were able to remove the evidence information completely since the blob detector in our case never failed. However, there might be case that a low threshold blob detector may fail to detect the foreground. In the next experiment, we explore how different values of ω accommodate this failure.

Experiment 3 (*effect of failure window on privacy loss and visual distortion*). In this experiment, our objective is to find the value of failure window (ω) for which the visual distortion is minimum for a near-zero privacy loss. We perform blurring and pixelization operations globally on ω consecutive frames in the video after the blob detection fails. Experiment is done with four values of ω: 0, 5, 10, and 15. The result is shown in Figure 15. As can be seen in the figure that for the given video (Video 3 in this case) and the blob detector used in our experiment, at a value of $\omega = 10$ we obtain a near-zero privacy loss. With this value of ω, the distortion is less than 0.2 with both blurring and pixelization operations (although pixelization causes slightly more distortion than blurring). We have conducted this experiment only for Video 3 because only this video had such a failure of the blob detector.

Experiment 4 (*when the blob detector fails*). The solution provided in the previous experiment fails in situations, where one person is detected and other could not be detected. The proposed method will only remove one person's identity. The other person will be left untransformed, and hence it might cause privacy loss. In this experiment we simulate such scenario by considering only the biggest blob detected in the video. Other blobs are assumed to be not detected. To improve the privacy loss in such scenarios, we use a very pessimistic approach of data transformation. Here we assume that someone is always there in the video and do global transformation when no blobs are detected. When the blob is detected, we do the space variant blurring according to the previous method; however, the image area outside estimated evidence region is globally transformed; unlike previous method where it is left unprocessed. Experiments show that this method performs better than global operations, but the perceptual quality is poorer in comparison to the normal space variant transformation discussed in previous experiments.

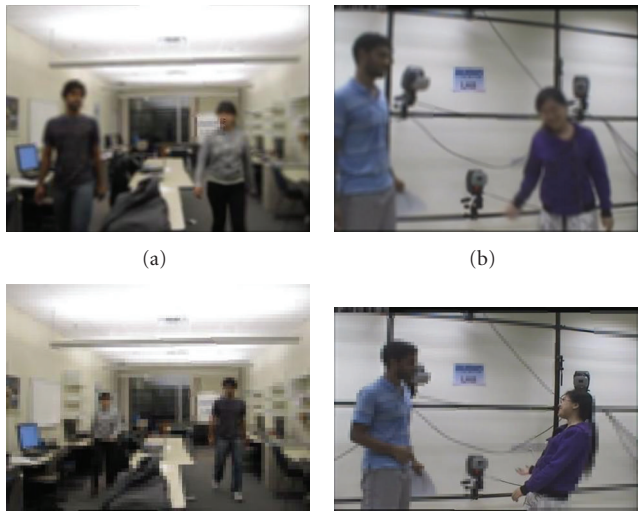


Figure 16: Output images from the pessimistic approach to overcome the failure of blob detector.

Figure 16 shows the output images for video 3 from the proposed algorithm. We notice that from Figures 16(a) to 16(c), we are able to hide the faces effectively even without global transformation. However, the method's success depends on the scenario and needs fine tuning. For example, in Figure 16(d) the other person's face is visible as it happens to fall in the outermost circle of the space variantly transformed first blob. Proper selection of the radius depends on the context and is out of scope of this paper. The qualitative results of the methods are provided in Figures 17 and 18. The figure can be compared with Figures 8 and 12 to conclude that the resulting video quality is worse than the normal space variant blurring is better than global transformation.

Experiment 5 (*validation with real surveillance data*). The five videos used in previous experiments cover various scenarios. The conclusions made for these scenarios are further validated by running the proposed method on real surveillance

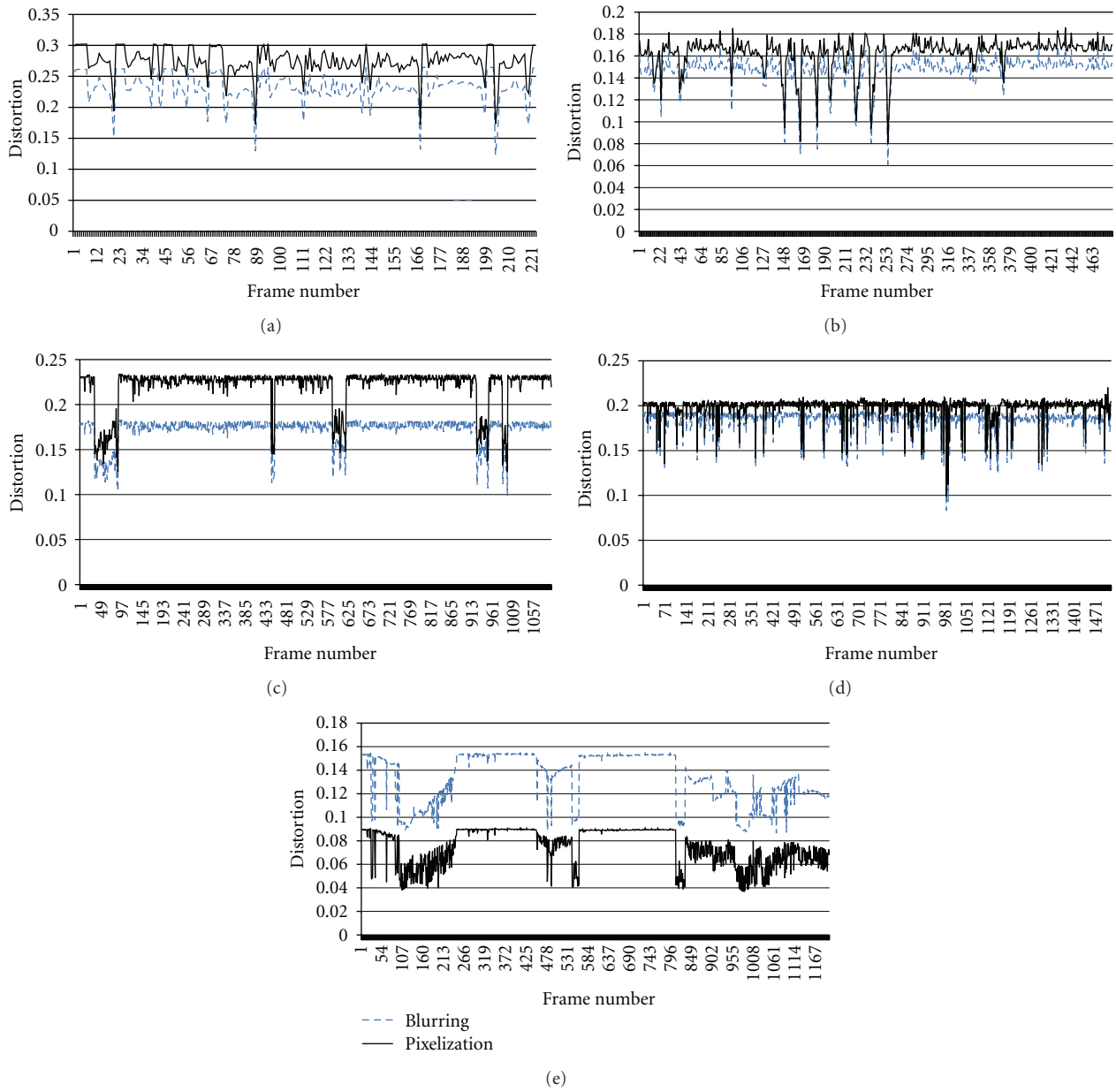


Figure 17: The distortion measures for blurring and pixelization for pessimistic approach.

footage of 24 hours, recorded at The University of Winnipeg. Since the video consists of 28216 frames, we omit the detailed distribution of the distortion values and provide the mean distortion in Figure 19 for the global method, the proposed method, and the pessimistic approach described in Experiment 4. A globally transformed background image of the video is shown in Figure 20. Other resulting figures are not shown due to privacy concerns.

We find that the results for real data are in agreement with our earlier conclusions. The proposed method causes less distortion than global transformation (63%) while the pessimistic approach causes more distortion than the proposed approach, though less than the global method. Further, the distortion caused by blurring is 22% less than that of pixelization.

4.2. Further Discussion. The main goal of this paper is to introduce adaptive transformation in spatial as well as temporal domains to overcome inaccuracies of detectors and to achieve more robust privacy. To the best of our knowledge, this is the first attempt towards reliable privacy with unreliable detectors. It is important to note that a tracking based solution could also be used for temporal adaption; however, it would again be limited by the accuracy of the tracker. Also, in real scenarios, it is very difficult to initialize the tracking with a generic template, and the tracker fails as

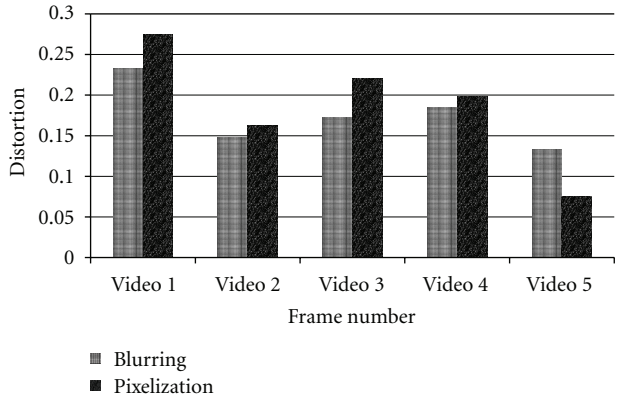


Figure 18: Mean distortion measures for blurring and pixelization for pessimistic approach.

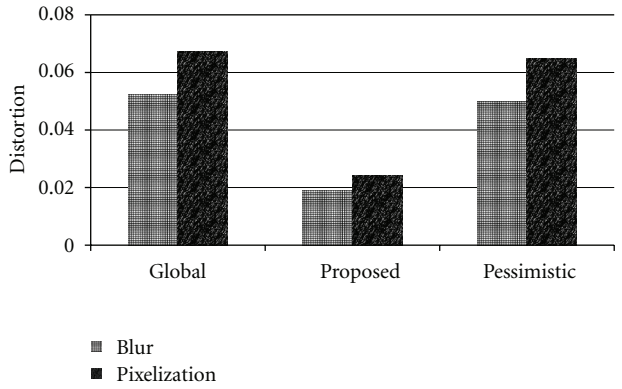


Figure 19: Mean distortion measures for blurring and pixelization for real surveillance data.

Figure 20: A globally transformed frame of real surveillance video.

soon as the person changes the posture. Therefore, we think that the proposed adaptive method is more robust.

5. Conclusions

The proposed adaptive approach of data transformation intelligently hides the evidence information in the video without much compromise with quality. It also provides robust privacy despite the inaccuracies of the detectors. Experimental results are provided to support our claims. For the CCTV scenario, we explored the adaptive transformation method to capitalize on the benefits of a global transformation while adapting it with the output of unreliable detectors. The following are the important conclusions of the paper:

(i) pixelization is found to be better than blurring and quantization for transforming static evidence regions with 8% less distortion than blurring and 55% less distortion than quantization;

(ii) the proposed method is more reliable than the selective obfuscation based methods and has 38% lesser visual distortion than global transformation;

(iii) for foreground transformation using space variant operations, blurring provides 11% less distortion than pixelization.

In the future, we want to deploy these methods in real implementations and perform a user study-based evaluation of privacy loss and distortion. It would be interesting to know how much distortion is acceptable to maintain a desired surveillance quality. Also, we want to extend the work by modeling the failure pattern of the detectors for the scenarios with more dynamic background and foreground.

Acknowledgment

Dr. P. K. Atrey's research contribution was supported by the Natural Sciences and Engineering Research Council of Canada.

References

[1] M. Boyle, C. Edwards, and S. Greenberg, "The effects of filtered video on awareness and privacy," in *Proceedings of the ACM Conference on Computer Supported Cooperative Work*, pp. 1–10, December 2000.

[2] M. Saini, P. K. Atrey, S. Mehrotra, S. Emmanuel, and M. Kankanhalli, "Privacy modeling for video data publication," in *Proceedings of the IEEE International Conference on Multimedia and Expo (ICME '10)*, pp. 60–65, July 2010.

[3] A. Senior, S. Pankanti, A. Hampapur et al., "Enabling video privacy through computer vision," *IEEE Security and Privacy*, vol. 3, no. 3, pp. 50–57, 2005.

[4] 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 *Proceedings of the 2nd ACM International Workshop on Video Sureveillance and Sensor Networks (VSSN '04)*, pp. 46–53, 2004.

[5] J. Wickramasuriya, M. Datt, S. Mehrotra, and N. Venkatasubramanian, "Privacy protecting data collection in media spaces," in *Proceedings of the 12th ACM International Conference on Multimedia*, pp. 48–55, usa, October 2004.

[6] T. Koshimizu, T. Toriyama, and N. Babaguchi, "Factors on the sense of privacy in video surveillance," in *Proceedings of the 3rd ACM Workshop on Continuous Archival and Retrievalof Personal Experiences (CARPE '06)*, pp. 35–43, 2006.

[7] B. Thuraisingham, G. Lavee, E. Bertino, J. Fan, and L. Khan, "Access control, confidentiality and privacy for video surveillance databases," in *Proceedings of the 11th ACM Symposium*

on Access Control Models and Technologies (SACMAT '06), pp. 1–10, June 2006.

[8] P. Carrillo, H. Kalva, and S. Magliveras, "Compression independent object encryption for ensuring privacy in video surveillance," in *Proceedings of the IEEE International Conference on Multimedia and Expo (ICME '08)*, pp. 273–276, June 2008.

[9] J. K. Paruchuri, S. C. S. Cheung, and M. W. Hail, "Video data hiding for managing privacy information in surveillance systems," *Eurasip Journal on Information Security*, vol. 2009, Article ID 236139, 7 pages, 2009.

[10] F. Z. Qureshi, "Object-video streams for preserving privacy in video surveillance," in *Proceedings of the 6th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS '09)*, pp. 442–447, 2009.

[11] 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*, vol. 5, no. 2, pp. 1–29, 2008.

[12] T. Spindler, C. Wartmann, and L. Hovestadt, "Privacy in video surveilled areas," in *Proceedings of the ACM International Conference on Privacy, Security and Trust*, pp. 1–10, 2006.

[13] A. Elgammal, R. Duraiswami, D. Harwood, and L. S. Davis, "Background and foreground modeling using nonparametric kernel density estimation for visual surveillance," *Proceedings of the IEEE*, vol. 90, no. 7, pp. 1151–1163, 2002.

[14] H. Kruegle, *CCTV Surveillance: Analog and Digital Video Practices and Technology*, Butterworth-Heinemann, Boston, Mass, USA, 2006.

[15] R. Kasturi, D. Goldgof, P. Soundararajan et al., "Framework for performance evaluation of face, text, and vehicle detection and tracking in video: data, metrics, and protocol," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 2, pp. 319–336, 2009.

[16] E. Hjelmås and B. K. Low, "Face detection: a survey," *Computer Vision and Image Understanding*, vol. 83, no. 3, pp. 236–274, 2001.

[17] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

[18] S. Chikkerur, V. Sundaram, M. Reisslein, and L. J. Karam, "Objective video quality assessment methods: a classification, review, and performance comparison," *IEEE Transactions on Broadcasting*, vol. 57, no. 2, pp. 165–182, 2011.

[19] K. Seshadrinathan, R. Soundararajan, A. C. Bovik, and L. K. Cormack, "Study of subjective and objective quality assessment of video," *IEEE Transactions on Image Processing*, vol. 19, no. 6, Article ID 5404314, pp. 1427–1441, 2010.

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

[21] C. Stauffer and W. E. L. Grimson, "Adaptive background mixture models for real-time tracking," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '99)*, vol. 2, pp. 246–252, June 1999.