Video Scrambling for Privacy Protection in Video Surveillance – Recent Results and Validation Framework

Frederic Dufaux

Laboratoire de Traitement et Communication de l'Information – UMR 5141 CNRS
Télécom ParisTech
F-75634 Paris Cedex 13, France
frederic.dufaux@telecom-paristech.fr

ABSTRACT

The issue of privacy in video surveillance has drawn a lot of interest lately. However, thorough performance analysis and validation is still lacking, especially regarding the fulfillment of privacy-related requirements. In this paper, we first review recent Privacy Enabling Technologies (PET). Next, we discuss pertinent evaluation criteria for effective privacy protection. We then put forward a framework to assess the capacity of PET solutions to hide distinguishing facial information and to conceal identity. We conduct comprehensive and rigorous experiments to evaluate the performance of face recognition algorithms applied to images altered by PET. Results show the ineffectiveness of naïve PET such as pixelization and blur. Conversely, they demonstrate the effectiveness of more sophisticated scrambling techniques to foil face recognition.

Keywords: Video surveillance, privacy, scrambling, face recognition

1. INTRODUCTION

Privacy is a cornerstone of our civilization and is essential in many societal functions [1]. However, this fundamental principle is quickly eroding due to widespread intrusiveness enabled by some modern information technologies.

In particular, privacy is quickly becoming a very central issue in video surveillance. While video surveillance can help repress crime and terrorism, hence benefiting society, the widespread use of security cameras has led to well documented forms of abuse, including: criminal abuse by law enforcement officers, institutional abuse by spying upon and harassing political activists, abuse for personal purpose such as stalking women or estranged girlfriends/spouses, discrimination including racial discrimination, voyeurism where bored male operators spy on women, and release of public camera footage in the public domain. Moreover, its big brother nature is hindering wider acceptance of video surveillance. The perspective of forthcoming powerful video analytics technologies, combined with pervasive networks of high resolution cameras is further raising the threat of privacy loss.

Fortunately, recent research results have shown that new Privacy Enabling Technologies (PET) are emerging with the potential to effectively protect privacy, without hampering video surveillance tasks. These results challenge the common conjecture that increased security should incur a loss in privacy. Recent overviews of PET are given in [2][3].

However, although the issue of privacy protection has drawn a lot of interest, thorough performance analysis is still lacking. In particular, it is paramount to validate proposed PET against user and system requirements for privacy. Moreover, it is still unclear whether current approaches can be efficiently integrated into existing surveillance architecture and deployed in large scale systems.

In this paper, we address the problem of assessing and validating PET. For this purpose, we identify five key evaluation criteria for effective PET: intelligibility of the video, cryptographic security, compression efficiency, computational complexity, and ease of integration in existing video surveillance systems.

The first criterion represents a significant challenge and is the focus of this study. More specifically, PET should make regions containing privacy-sensitive information unintelligible. Simultaneously, the remaining of the scene should be intelligible in order not to hamper video surveillance tasks. However, assessing the intelligibility/unintelligibility of a video is a difficult problem. Existing objective visual quality measures, such as Peak-Signal-to-Noise-Ratio (PSNR) and Structural Similarity (SSIM) index [4], have been designed with the goal to assess distortions resulting from typical image and video processing techniques, but not to address intelligibility.

In this paper, we tackle this problem by assessing the capability of PET to make facial information unintelligible and henceforth to foil face recognition techniques and conceal identity. Indeed, this is a major threat to privacy in video surveillance. A face de-identification algorithm is proposed in [5], which preserves many facial characteristics but makes the face unrecognizable. It is also shown that simple PET do not prevent successful face recognition. In our previous work [6], we defined a framework to evaluate the performance of face recognition algorithms applied to images altered by PET, based on the Face Identification Evaluation System (FIES) [7]. Experiments on the Facial Recognition Technology (FERET) database [8] showed the ineffectiveness of naïve PET such as pixelization and blur, and demonstrated the effectiveness of more sophisticated scrambling techniques to foil face recognition. In this paper, we extend this earlier work. We identify and discuss evaluation criteria for PET. We also conduct more extensive experiments, including with PSNR and SSIM objective quality measures.

This paper is structured as follow. An overview of recent PET is presented in Sec. 2. Evaluation criteria to assess PET effectiveness are presented in Sec. 3. Next, a framework for face identification evaluation is presented in Sec. 4. An outline of four PET under consideration is given in Sec. 5. In order to validate PET, performance assessment using the proposed framework is analyzed in Sec. 6. Finally, conclusions are summarized in Sec. 7.

2. PRIVACY ENABLING TECHNOLOGIES

The raising awareness about privacy issues in surveillance systems has led to the development of new PET with the goal to effectively protect privacy [2][3].

The system introduced in [9] relies on computer vision to analyze the video content and to automatically extract its components. Different users can selectively get access to these components, depending on their access-control rights. More specifically, the system renders a different version of the video where privacy-sensitive objects have been hidden. This is achieved while information required to fulfill the surveillance task is preserved. The paper also describes a PrivacyCam, with built-in privacy protection tools, which directly outputs video streams with privacy-sensitive information removed.

The Networked Sensor Tapestry (NeST) architecture proposed in [10] supports secure capture, processing, sharing and archiving of surveillance data. It relies on privacy filters which operate on incoming video sensor data to remove privacy-sensitive information. These filters are specified using a privacy grammar.

With Privacy through Invertible Cryptographic Obscuration (PICO) proposed in [11], data corresponding to faces is encrypted in order to conceal identity. The process is reversible for authorized users in possession of the secret encryption key. In other words, it does not undermine the objective of surveillance, as a subject can still be identified by decrypting the face, provided an appropriate warrant is issued. Similarly, a permutation-based encryption technique in the pixel domain is introduced in [12]. The solution remains independent of the compression algorithm and is robust to transcoding. The scheme presented in [13], Secure Shape and Texture SPIHT (SecST-SPIHT), permits secure coding of arbitrarily shaped visual objects. More specifically, a novel selective encryption is introduced, applied in the compressed domain. Likewise, data hiding method based on chaos cryptography is introduced in [14]. The technique is applied to selected Regions of Interest (ROI) corresponding to privacy-sensitive information, and allows for several levels of concealment.

The methods in [15][16] propose PET for JPEG 2000 video [17]. Conditional access control techniques are proposed in [15] to scramble ROIs, e.g. corresponding to people or faces. The scrambling is applied either in wavelet-domain or codestream-domain. In [16], code-blocks corresponding to ROI are trimmed down to the lowest quality layer of the codestream. Subsequently, the quality of the ROI can be decreased by limiting the video bit rate.

Two efficient region-based transform-domain and codestream-domain scrambling techniques are proposed in [18] to hide privacy-sensitive information in MPEG-4 video [19]. In the first approach, the sign of selected transform coefficients is pseudo-randomly inverted during encoding. In the second approach, bits of the codestream are pseudo-randomly flipped after encoding. In [20], the region-based transform-domain scrambling is extended to H.264/AVC [21]. In particular, to discriminate between scrambled and unscrambled regions, the technique exploits the Flexible Macroblock Ordering (FMO) mechanism of H.264/AVC to define two slice groups composed of MacroBlocks (MB) corresponding to the foreground and background respectively.

The technique in [22] removes privacy-sensitive information from the video sequence. A perceptually-based compressed-domain watermarking technique is then used to securely embed this data in the video stream. Similarly, a secure reversible data hiding technique is introduced in [23] for privacy data embedding. A framework for privacy data management is also proposed to allow individual users to control access to their private data.

Face recognition techniques pose the risk to automatically and quickly identify people captured by a video surveillance system. This issue is addressed in [5], where a de-identifying algorithm is introduced to effectively foil face recognition. A face anonymization framework for mobile phones is proposed in [24]. In this system, people who do not want to have their picture taken inform other mobile phone users in the vicinity using Bluetooth. Whenever a picture is taken, the corresponding faces are then anonymized.

3. EVALUATION CRITERIA

Despite the significant efforts to develop new PET in recent years, a methodical validation of their effectiveness is still lacking. More specifically, it is important to carefully assess these solutions against user and system requirements for privacy. In this section, we discuss evaluation criteria for PET.

3.1. Intelligibility / Unintelligibility

The foremost criterion for PET is to successfully conceal private information in the video stream. It typically implies to render some regions, with privacy-sensitive data, unintelligible. At the same time, PET should leave the remaining of the video scene comprehensible in order not to hamper surveillance tasks.

In most data security applications, the objective is to guarantee full confidentiality of the message. It is usually achieved by applying encryption techniques. In contrast, in many multimedia security applications, it is sufficient to partially protect the data so that versions of the content with a quality above a commercially valuable threshold are protected, but low quality/resolution previews remain clear. In this case, selective encryption techniques are most appropriate [25].

In the context of PET for video surveillance, the situation is to a certain extent similar. More specifically, PET should hide or deteriorate the visual quality of the scene regions corresponding to privacy-sensitive data. To be effective, the alteration introduced should be such that it prevents identification. Simultaneously, the distortion should not prevent an operator to correctly interpret the scene during surveillance operations. The key challenge is to assess whether this dual requirement is fulfilled.

Existing objective visual quality measures, such as SSIM [4], have been essentially developed to evaluate common image/video coding and processing techniques. Furthermore, they assess subjective quality instead of intelligibility which is the primary concern with PET. Finally, they have been tuned in a high quality range, rather than the highly altered range typical with PET.

In [26], two visual similarity measures are proposed in order to measure security for multimedia encryption. The first one is based on luminance similarity, whereas the second one considers edge similarity. However, none of these measures address the intelligibility of the content and hence the ability to conceal private data.

A model to assess privacy loss is introduced in [27], based on an analogy with statistical databases. It is shown that privacy issues encompass both explicit and implicit inference channels. The latter considers the situation when the identity of an individual can be indirectly deduced from the examination of the video content, e.g. based on location, time or behavior.

One of the major privacy threats in video surveillance is to automatically identify people in the scene using face recognition techniques. Hence, the effectiveness of face recognition techniques on images altered by PET can be used as a validation criterion to assess the unintelligibility of private information and hence the usefulness of PET. In [5], it is shown that simple ad-hoc de-identification methods do not prevent successful face recognition. To successfully tackle this issue, a more sophisticated algorithm to de-identify faces is required, such that many facial characteristics are preserved but the face cannot be reliably recognized. In [6], a framework is proposed to validate PET. More specifically, this framework, based on the FIES [7], assesses the capability to effectively conceal identity by evaluating the performance of face recognition algorithms on images altered by PET.

3.2. Cryptographic Security

Many PET rely on cryptographic techniques in order to obscure regions containing privacy-sensitive data, for instance [11][12][13][14][15][18][20]. Cryptanalysis of a system is often evaluated under the assumption that the objective is to recover the whole encrypted message. In this context, the strength of the protection technique corresponds to the difficulty of finding the secret encryption key.

However, the above scenario has a few shortcoming, as shown in [28]. Firstly, information leakage occurs. Namely, part of the data which is not encrypted can be used in order to guess/interpolate the encrypted part. Secondly, video content typically presents well known statistical and structural properties which can be exploited by an attacker. Finally, even though an attacker cannot totally recover the protected data, the security is still considered as compromised if he is capable of recovering an image with sufficiently improved subjective quality. For PET, the threshold is whether the visual quality is sufficient to identify private information.

3.3. Compression Efficiency

Bandwidth is a precious resources in video surveillance systems. On the one hand, a class of PET approaches relies on encryption or scrambling [11][12][13][14][15][18][20]. In this way, the statistics of the data is drastically altered. Depending on how it is integrated within a compression scheme, it may significantly increase bit rate requirements. On the other hand, another class of techniques such as [22][23] embed privacy-sensitive information using data hiding. Straightforwardly, it may result in sizeable data overhead.

To be effective, PET should have a minimal impact on compression efficiency.

3.4. Computational Complexity

Computational complexity is also an important issue in video surveillance application. Indeed, extra complexity, either embedded directly in the camera or performed on a server, has a direct impact on the hardware cost. In particular, cryptographic functions tend to entail significant computations. For instance, one of the major interests of selective encryption techniques is to reduce complexity requirements by processing a subset of the data only.

3.5. Integration in Existing Surveillance Architecture / Utility for Surveillance

Easy integration of PET in existing video surveillance infrastructure is another important criterion in order to foster rapid adoption of the technology. Compatibility with legacy systems ensure broader and cost-effective applicability of PET. Accordingly, approaches which rely on widely used video coding standards (e.g. H.264/AVC, MPEG-4, Motion JPEG), instead of proprietary representations, should be preferred. In addition, PET which preserve syntax format compliance offer a considerable advantage. In this case, standard decoders can correctly decode and display the video stream, although some regions may be concealed. Moreover, preserving the stream syntax and its features enables content adaptation based on scalability or transcoding during network transmission.

Another valuable feature is to transmit the same protected video stream to all end-users, regardless of their credentials.

Finally, video surveillance is most often used in postmortem forensic analysis by law enforcement authorities. For this purpose, it is paramount that PET are fully reversible. Namely, it should be possible, for authorized users, to recover the unaltered privacy-sensitive information. Obviously, some trivial PET approaches merely applying blur, noise, or black box obscuration to hide private data do not fulfill this important requirement.

4. FRAMEWORK FOR FACE IDENTIFICATION EVALUATION

The objective of this paper is to validate the anonymity functionality of PET. For this purpose, we use FIES [7], which provides standard face recognition algorithms and standard statistical methods for assessing performances. A brief description of the framework is given hereafter.

4.1. Face Recognition

We consider two face recognition algorithms, namely, Principal Components Analysis (PCA) [29] and Linear Discriminant Analysis (LDA) [30].

In PCA, eigenfaces corresponding to the eigenvectors of the covariance matrix of training face examples are computed. Face images are then projected onto the eigenfaces basis. In other words, a linear transformation is applied to rotate feature vectors from the initially large and highly correlated subspace to a smaller and uncorrelated subspace. Distance between pair of images can then be computed in the eigenfaces subspace. Hereafter, the distance between the feature vectors, u and v, is given by the Euclidian measure

$$D_{\text{Euclidian}}(u,v) = \sqrt{\sum_i (u_i - v_i)^2} \tag{6}$$

PCA has shown to be effective for face recognition. Firstly, it can be used to reduce the dimensionality of the feature space. Secondly, it eliminates statistical covariance in the transformed feature space. In other words, the covariance matrix for the transformed feature vectors is always diagonal.

LDA aims at finding a linear transformation which stresses differences between classes while lessening differences within classes, where a class corresponds to all images of a given individual. The resulting transformed subspace is linearly separable between classes. In [30], PCA is first performed to reduce the feature space dimensionality. LDA is then applied to further decrease the dimensionality while safeguarding the distinctive characteristics of the classes. The final subspace is obtained by multiplying the PCA and LDA basis vectors. Feature vectors, i.e. face images are then projected onto these basis vectors. Finally, the soft distance proposed in [30] is used, namely

$$D_{\text{LDAsoft}}(u,v) = \sum_i \lambda_i^{0.2} (u_i - v_i)^2 \tag{7}$$

where is λ_i is the corresponding eigenvalue.

4.2. Face Identification and Evaluation System

FIES is composed of four main components: image pre-processing, training, testing and performance analysis [7]. The framework is illustrated in Figure 1.

The preprocessing step aims at reducing detrimental variations between images. Faces are firstly geometrically normalized by aligning the eye coordinates. Then, an elliptical mask is used to crop the images. More specifically, the face region between the forehead and chin as well as left cheek and right cheek is retained, while the remaining is discarded. Finally, histogram equalization is performed, and contrast and brightness are normalized.

The next step is training. Its purpose is to create the subspace into which test images are subsequently projected and matched. In this paper, we consider PCA and LDA techniques as previously described. Training is performed using a training set of images.

In the testing step, a distance matrix is computed in the transformed subspace for all test images. In our experiments, we use a Euclidian distance for PCA and the soft distance for LDA, as defined in Eq. (6) and (7). At this stage, two image sets are defined: the gallery set is made of known faces, whereas the probe set corresponds to faces to be recognized.

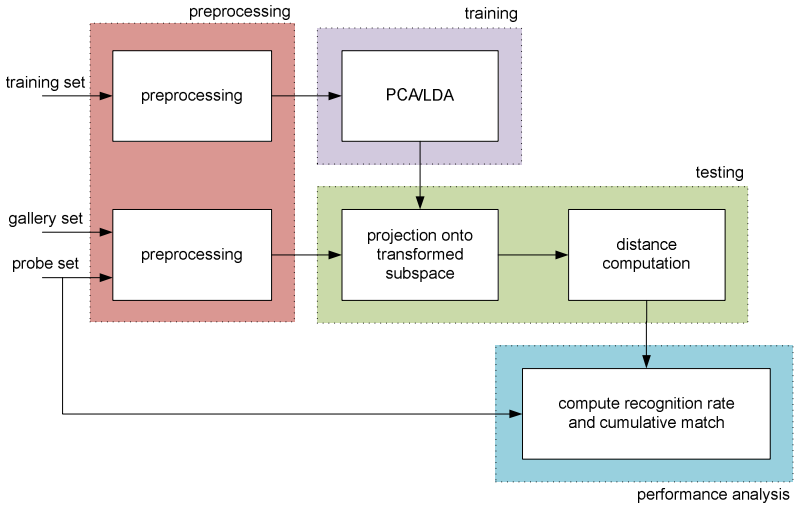

Figure 1 – Framework for face identification and evaluation.

Finally, face recognition performance is analyzed. More specifically, a cumulative match curve is generated. For this purpose, for each probe image, the recognition rank is computed. Namely, a rank 0 means that the best match is of the same subject, a rank 1 means that the best match is from another person but the second best match is of the same subject, etc. Then, the cumulative match curve is obtained by summing the number of correct matches for each rank.

4.3. Attacks under Consideration

We study two types of attack. We consider a simple attack, referred to as Attack A, where training and gallery sets are made of unaltered images. Conversely, probe set corresponds to images with privacy protection. In other words, altered images are merely processed by the face recognition algorithms without taking into account the fact that PET have been applied.

We then consider a more sophisticated second attack, referred to as Attack B. Namely, PET are now applied to all images in the training, gallery and probe sets. This corresponds to an attacker which gets access to protected data. Alternatively, an attacker may attempt replicating the alteration due to PET on his own training and gallery sets.

Table 1 summarizes the characteristics of both attacks.

set	Attack A	Attack B
training	unaltered	privacy protection
gallery	unaltered	privacy protection
probe	privacy protection	privacy protection

Table 1 – Attacks under consideration.

5. PET UNDER CONSIDERATION

In this section, we briefly describe four PET that we will subsequently evaluate for their capability to hide facial information and to provide anonymity.

As reference, we first consider two naïve methods, applying simple pixelization or Gaussian blur. We also consider two more sophisticated ROI-based transform-domain scrambling methods [20]. Both methods are applied jointly with H.264/AVC encoding [21], which is becoming the prevalent format in video surveillance systems. The first method pseudo-randomly inverts the sign of transform coefficients of blocks belonging to ROI. The second one applies a pseudo-random permutation of the transform coefficients in blocks corresponding to ROI.

These four approaches to provide anonymity are detailed in the following subsections.

5.1. Pixelization

We first consider pixelization as a naïve approach for privacy protection. Pixelization consists in noticeably reducing resolution in ROI. In practice, it can be achieved by substituting a square block of pixels with its average. Explicitly, pixelization of the image $I(x,y)$ is given by

$$I_{\text{pixelization}}(x,y) = \frac{1}{b^2} \sum_{i=0,\dots b-1} \sum_{j=0,\dots b-1} I\left(\left\lfloor \frac{x}{b} \right\rfloor + i, \left\lfloor \frac{y}{b} \right\rfloor + j \right) \tag{1}$$

where x and y are the image pixel coordinates, b is the block size and $\lfloor \cdot \rfloor$ denotes the floored division.

Pixelization has the advantage to be very simple and easy to integrate in existing systems. Consequently, it is commonly used in television news and documentaries in order to obscure the faces of suspects, witnesses or bystanders to preserve their anonymity. The same technique is also used to censor nudity or to avoid unintentional product placement on television.

Straightforwardly, using pixelization, privacy-sensitive information is lost and the process is irreversible. Another drawback of this approach is that integrating pixels along trajectories over time may allow to partly recovering the concealed information.

5.2. Gaussian Blur

The second naïve approach for privacy protection removes details in ROI by applying a Gaussian low pass filter. More precisely, Gaussian blur is obtained by the convolution of the image $I(x,y)$ with a 2D Gaussian function $G(x,y)$

$$I_{\text{Gaussian blur}}(x,y) = I(x,y) * G(x,y) \tag{2}$$

with $G(x,y)$ defined by

$$G(x,y) = \frac{1}{2\pi\sigma^2} e^{-\frac{(x^2+y^2)}{2\sigma^2}} \tag{3}$$

where σ is the standard deviation. Again, the process is very simple and easy to implement. However, it is irreversible and privacy-sensitive information is irremediably lost. Blurring is sometimes preferred to pixelization in order to obscure privacy-sensitive information.

5.3. Scrambling by Random Sign Inversion

Next, we consider a ROI-based transform-domain scrambling method for H.264/AVC. First, the method scrambles the quantized transform coefficients of each 4x4 block of the ROI, corresponding to privacy-sensitive information, by pseudo-randomly flipping their sign [20]. More specifically, defining the vector of quantized transform coefficients qcoeff[i] with i=0..15, the scrambling consists in performing the following operation for each i

$$qcoeff[i] = \begin{cases} -qcoeff[i] & \text{if } random\_bit = 1 \\ +qcoeff[i] & \text{otherwise} \end{cases} \tag{4}$$

The scrambling process is driven by a Pseudo Random Number Generator (PRNG) which is initialized by a seed value. The seed is encrypted, e.g. using public key encryption, and embedded in the compressed stream as private date.

The process is illustrated in Figure 2.

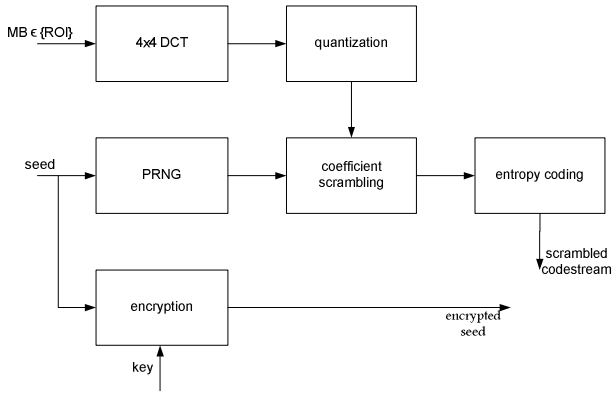

Figure 2 – Region-based transform-domain scrambling by random sign inversion or random permutation.

The method is fully reversible. Namely, authorized users, in possession of the secret encryption key, can reverse the scrambling process and recover the truthful scene. Conversely, other users obtain a video sequence where ROI have severe noise, concealing privacy-sensitive information.

Two slice groups are defined using FMO to distinguish between the scrambled ROI and the unscrambled background. In this way, background data will not use scrambled ROI data for spatial intra prediction. An added benefit of FMO is that the shape of the scrambled ROI is conveyed to the decoder which needs this information for unscrambling.

This method offers a number of advantages. The same scrambled stream is transmitted to all users independently from their access rights. Furthermore, the syntax of the compressed stream remains standard compliant. Hence, it can be used in existing video surveillance infrastructures. The scrambling is confined to ROI, whereas the background remains unaltered. Finally, it has a small impact in terms of coding efficiency, and requires a low computational complexity.

5.4. Scrambling by Random Permutation

Finally, we consider an alternative ROI-based transform-domain scrambling method for H.264/AVC. In this method, a random permutation is applied to rearrange the order of transform coefficients in 4x4 blocks corresponding to ROI [20]. The method is depicted in Figure 2. The random permutation is expressed as follow

$$\begin{pmatrix} 0 & 1 & \dots & 14 & 15 \\ x_0 & x_1 & \dots & x_{14} & x_{15} \end{pmatrix} \tag{5}$$

The Knuth shuffle is used to generate a permutation of n items with uniform random distribution [31]. More explicitly, it starts from the identity permutation and scans through each position i from 0 to 14, swapping the element currently at position i with the element at an arbitrarily chosen positions from i through 15.

The remaining of the algorithm is similar to the scrambling by random sign inversion as described in Sec. 5.3. Hence, this method provides the same characteristics and advantages.

5.5. Sample Images with Privacy Protection

The results of the four PET considered in this paper, namely pixelization (b={8, 16}), Gaussian blur (σ={8, 12}), scrambling by random sign inversion and scrambling by random permutation are illustrated in Figure 3 for a sample image of the FERET database.

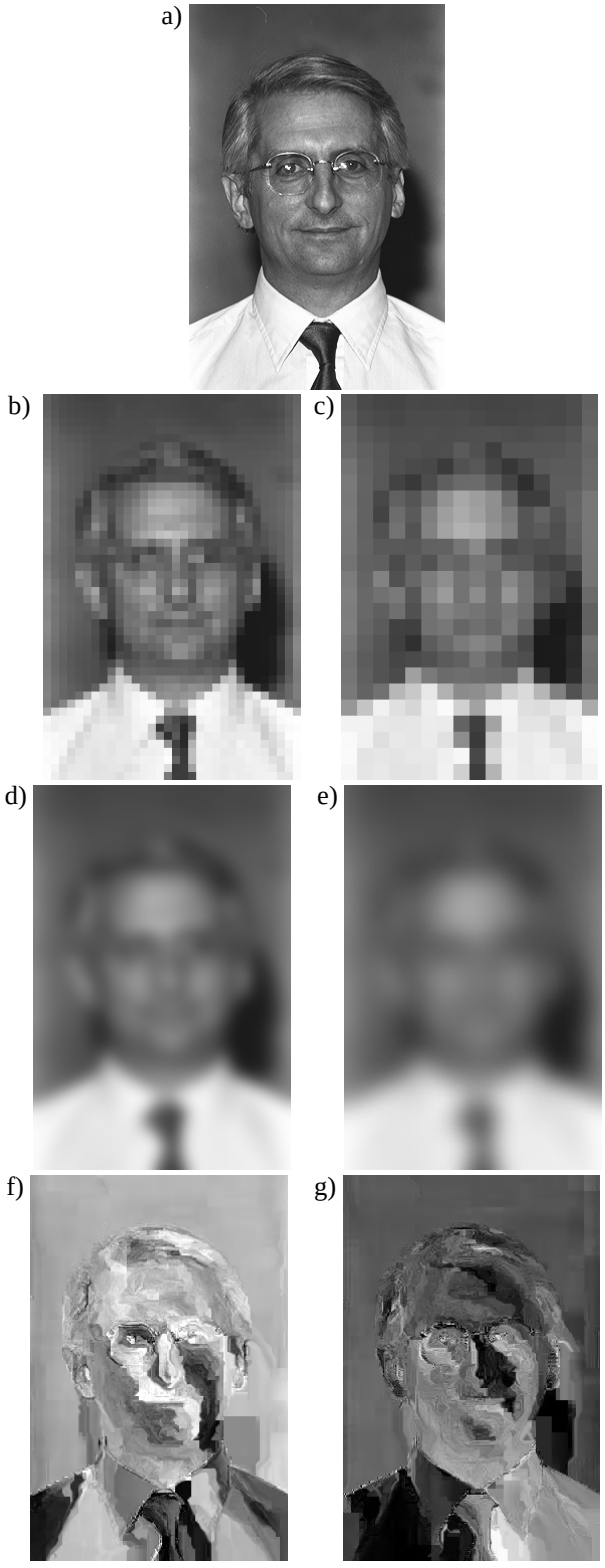

Figure 3 – Examples of privacy protection approaches: a) original image, b) pixelization with b=8, c) pixelization with b=16, d) Gaussian blur with σ=8, e) Gaussian blur with σ=12, f) scrambling by random sign inversion, g) scrambling by random permutation.

6. PRIVACY ENABLING TECHNOLOGIES PERFORMANCE ASSESSMENT RESULTS

We now describe experiments carried out in order to assess PET. Results are then reported and analyzed.

6.1. Test Data

In this paper, we use the grayscale FERET database [8] to carry out experiments. Indeed, this database is widely used for face recognition research, although it is not representative of typical video surveillance footage. From this database, we consider a subset of 3368 images of frontal faces for which eye coordinates are available. The images have 256 by 384 pixels with eight-bit per pixel. We further consider two series of images denoted by 'fa' and 'fb'. The 'fa' indicates a regular frontal image, and the 'fb' indicates an alternative frontal image, taken within seconds of the corresponding 'fa' image, where a different facial expression was requested from the subject.

In our experiments, we use standard training, gallery and probe sets from the FERET test. More specifically, the training set includes 501 images from the 'fa' series. In turn, the gallery set is composed of 1196 from the 'fa' series, whereas the probe set is made of 1195 images from the 'fb' series.

6.2. Objective Image Quality Measures

Next, we assess the quality of images altered by the four PET under consideration using objective image quality measures. To measure the similarity between the altered and unaltered images, we use PSNR

$$\mathrm{PSNR}(I,K) = 10\log_{10}\left(\frac{255^2}{\frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}\left(I(i,j)-K(i,j)\right)^2}\right) \tag{8}$$

where I and K are the processed and reference images respectively, with a size of m x n. However, PSNR is often weakly correlated with human perception.

Therefore, we use a second measure, the perceptually-based SSIM index [4], which assesses the degradation of structural information. SSIM, computed on windows of an image, is given by

$$\mathrm{SSIM}(I_w,K_w) = \frac{\left(2\mu_I\mu_K + C_1\right)\left(2\sigma_{IK} + C_2\right)}{\left(\mu_I{}^2 + \mu_K{}^2 + C_1\right)\left(\sigma_I{}^2 + \sigma_K{}^2 + C_2\right)} \tag{9}$$

where I_w and K_w are two windows in the altered and unaltered images respectively, μ_I and μ_K are the average of I_w and K_w, $\sigma_I{}^2$ and $\sigma_K{}^2$ are the variances of I_w and K_w, and C_1 and C_2 are two constants to avoid instability. SSIM takes value in the interval [-1, 1], where SSIM=1 indicates that both processed and reference images are identical.

Table 2 shows the PSNR and SSIM values obtained using the four PET: pixelization, Gaussian blur, scrambling by sign inversion and scrambling by permutation. The reported values correspond to the average over all 3386 images in the considered FERET subset.

PET	PSNR	SSIM
Pixelization b=8	24.14	0.71
Pixelization b=16	21.11	0.64
Gaussian blur σ=8	22.56	0.71
Gaussian blur σ=12	20.78	0.68
Scrambling sign inversion	8.47	0.42
Scrambling permutation	9.09	0.47

Table 2 – Objective quality measures.

These figures are very instructive. They clearly show that all the PET under consideration lead to low subjective quality.

Straightforwardly, the image quality decreases when the block size increases in pixelization, or when σ increases in Gaussian blur. Nevertheless, for these two naïve PET, PSNR remains above 20 dB, whereas SSIM stays above 0.6. Conversely, both scrambling approaches result in significantly higher distortions when compared to pixelization or Gaussian blur, with PSNR below 10 dB and SSIM below 0.5. However, these results do not allow to assess the intelligibility of the concealed data, and therefore are not sufficient to validate the effectiveness of PET.

6.3. Face Recognition Performance Analysis

We now evaluate the capacity of PET to hide distinguishing facial information in order to foil face recognition techniques and hence to conceal the identity of a person.

In the first round of experiments, we consider the simple Attack A (see Sec. 4.3). Figure 4 and Figure 5 show cumulative match curves for PCA and LDA respectively, comparing the recognition rate as a function of the rank for original image data as well as for the four considered PET.

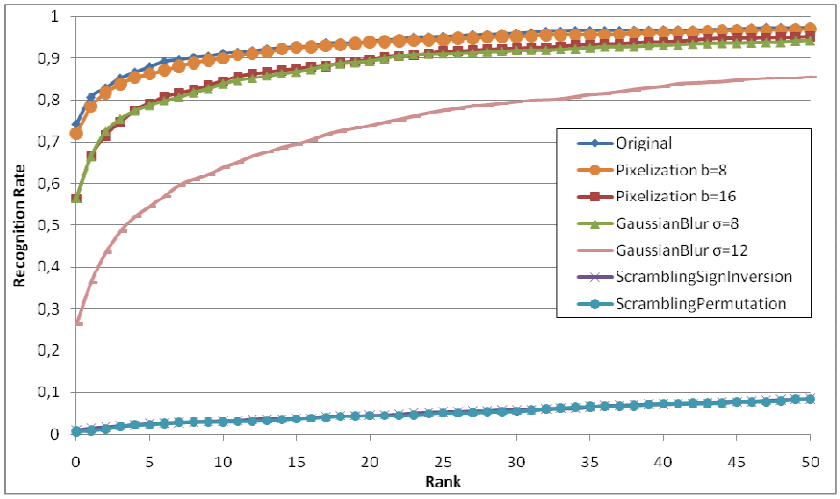

Figure 4 – Cumulative match curve for PCA with Euclidian distance: performance comparison for Attack A.

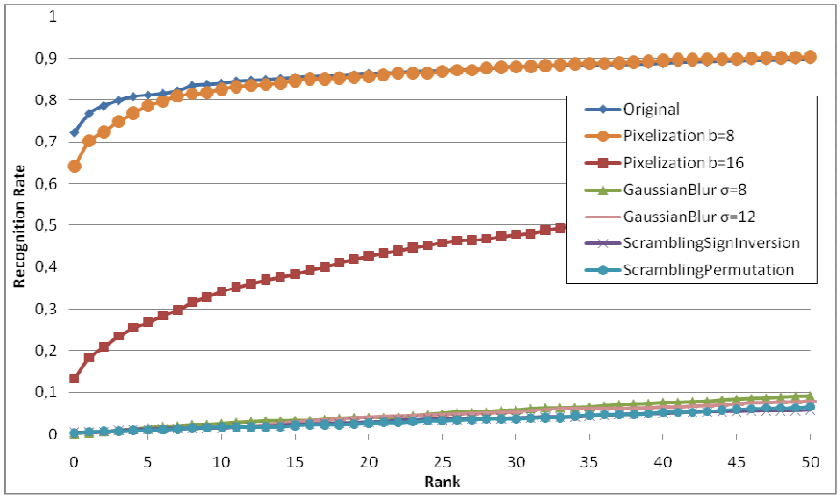

Figure 5 – Cumulative match curve for LDA with soft distance: performance comparison for Attack A.

It can be observed that for both PCA and LDA schemes applied on original images, recognition rate is superior to 70% at rank 0 (i.e. the best match is of the same subject as the probe), and superior to 90% at rank 50.

When applying a Gaussian blur, the performance drops radically for LDA. However, recognition rate remains high for PCA with 56% (σ=8) and 26% (σ=12) success at rank 0. Pixelization fares worse at hiding identity. With *b*=8, the performance only drops marginally compared to the recognition rate on original images. It is not as bad with *b*=16, but the recognition rate remains high with 56% and 13% at rank 0 for PCA and LDA respectively.

However, results clearly show that both region-based transform-domain scrambling approaches are successful at hiding identity. The recognition rate is nearly 0% at rank 0, and remains below 10% at rank 50, for both PCA and LDA algorithms. In addition, it can be observed that both random sign inversion and random permutation schemes achieve nearly the same performance.

In the second round of experiments, we consider the more sophisticated Attack B (see Sec. 4.3). Figure 6 and Figure 7 show corresponding cumulative match curves for PCA and LDA respectively.

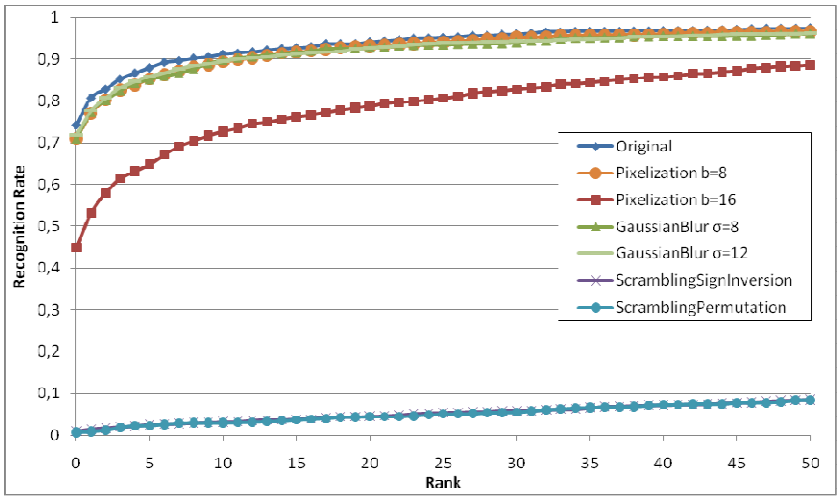

Figure 6 – Cumulative match curve for PCA with Euclidian distance: performance comparison for Attack B.

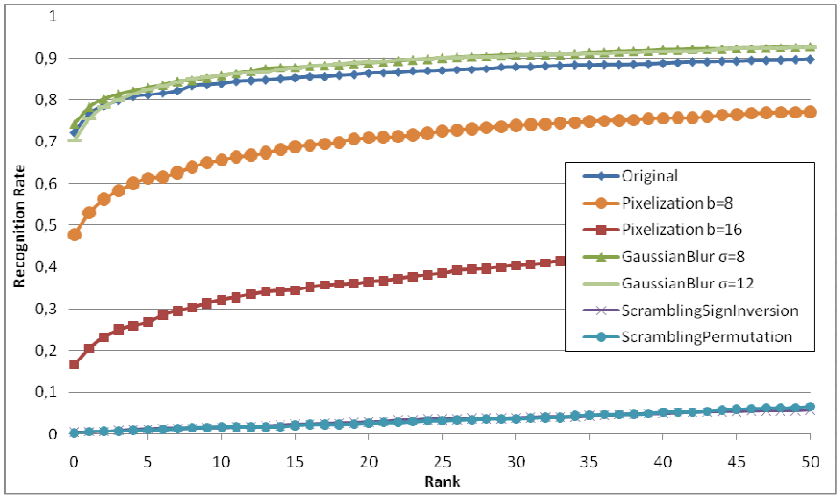

Figure 7 – Cumulative match curve for LDA with soft distance: performance comparison for Attack B.

With Gaussian blur, the performance remains nearly identical to the recognition rate on original data. It even improves slightly for LDA. Pixelization is not much better at hiding facial information. With b=16, the recognition rate is still 45% and 17% at rank 0 for PCA and LDA respectively.

Finally, both region-based transform-domain scrambling approaches are again successful at hiding identity. The recognition rate is nearly 0% at rank 0 for both PCA and LDA algorithms.

7. CONCLUSIONS

In this paper, we have considered the problem of validating PET for video surveillance applications. We have first reviewed some existing PET solutions. We have then identified relevant evaluation criteria as well as challenges for performance assessment. We have also described a framework to verify the effectiveness of PET at hiding distinguishing facial information and hence concealing identity.

We have conducted rigorous and comprehensive experiments on the FERET database using objective image quality metrics on the one hand, and PCA and LDA face recognition algorithms on the other hand. Results have shown that naïve PET approaches such as Gaussian blur or pixelization are ineffective at providing anonymity. In both cases, the recognition rate remains significant. Finally, we have shown that region-based transform-domain scrambling approaches are successful at hiding identity, with the recognition rate dropping to nearly 0%.

Future work will concentrate in further analyzing the performance of PET, verifying that they can successfully address privacy issues. In particular, it is important to carry out experiments on larger data sets and using more realistic video surveillance footage. It is also imperative to better understand user and system requirements regarding privacy issues. Finally, performance analysis should also include the impact on compression efficiency, complexity, and security against attacks.

REFERENCES

[1] M. Caloyannides, “Society Cannot Function Without Privacy”, IEEE Security and Privacy, vol. 1 , no. 3, pp. 84-86, May 2003.

[2] A. Cavallaro, “Privacy in Video Surveillance”, IEEE Signal Proc. Magazine, vol. 24, no. 2, pp. 168-169, March 2007.

[3] A. Senior, “Protecting Privacy in Video Surveillance” Springer, 2009.

[4] Z. Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli, “Image Quality Assessment: From Error Visibility to Structural Similarity”, IEEE Trans. on Image Processing, vol. 13, no. 4, pp. 600-612, April 2004.

[5] E. Newton, L. Sweeney, and B. Malin, “Preserving Privacy by De-identifying Face Images”, IEEE Trans. on Knowledge and Data Engineering, vol. 17, no. 2, pp. 232-243, February 2005.

[6] F. Dufaux and T. Ebrahimi, “A Framework for the Validation of Privacy Protection Solutions in Video Surveillance”, in Proc. IEEE International Conference on Multimedia & Expo (ICME 2010), Singapore, July 2010.

[7] Evaluation of face recognition algorithms web site, http://www.cs.colostate.edu/evalfacerec.

[8] The Facial Recognition Technology (FERET) database, http://www.itl.nist.gov/iad/humanid/feret/feret_master.html

[9] A.W. Senior, S. Pankanti, A. Hampapur, L. Brown, Y.-L. Tian, A. Ekin, J. Connell, C.F. Shu and M. Lu, “Enabling Video Privacy through Computer Vision”, IEEE Security and Privacy, vol. 3, no.3, pp. 50-57, May-June 2005.

[10] D. A. Fidaleo, H.-A. Nguyen, M. Trivedi, “The networked sensor tapestry (NeST): a privacy enhanced software architecture for interactive analysis of data in video-sensor networks”, Proc. of the ACM 2nd Int. Workshop on Video Surveillance & Sensor Networks, New York, NY, 2004.

[11] T.E. Boult, “PICO: Privacy through Invertible Cryptographic Obscuration”, IEEE Workshop on Computer Vision for Interactive and Intelligent Environments, Nov. 2005.

[12] P. Carrillo, H. Kalva and S. Magliveras, “Compression Independent Reversible Encryption for Privacy in Video Surveillance”, EURASIP Journal on Information Security, vol. 2009, Article ID 429581, doi:10.1155/2009/429581.

[13] K. Martin, and K.N. Plataniotis, “Privacy Protected Surveillance Using Secure Visual Object Coding”, IEEE Trans. on Circuits and Systems for Video Technology, vol. 18, no. 8, pp. 1152-1162, Aug. 2008.

[14] Sk. Md. Mizanur Rahman, M. A. Hossain, H. Mouftah, A. El Saddik, E. Okamoto, “A Real-Time Privacy-Sensitive Data Hiding Approach Based on Chaos Cryptography”, in Proc. IEEE International Conference on Multimedia & Expo (ICME 2010), Singapore, July 2010.
[15] F. Dufaux, and T. Ebrahimi, “Video Surveillance using JPEG 2000”, in SPIE Proc. Applications of Digital Image Processing XXVII, Denver, CO, Aug. 2004.
[16] I. Martinez Ponte, X. Desurmont, J. Meessen, and J.-F. Delaigle, “Robust Human Face Hiding Ensuring Privacy” in Proc. of International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS), Montreux, Switzerland, April 2005.
[17] D. Taubman and M. Marcellin, JPEG 2000: Image Compression Fundamentals,Standards and Practice. Norwell, MA: Kluwer, 2002.
[18] F. Dufaux and T. Ebrahimi, “Scrambling for Privacy Protection in Video Surveillance Systems”, IEEE Trans. on Circuits and Systems for Video Technology, vol. 18, no. 8, pp. 1168-1174, Aug. 2008.
[19] T. Ebrahimi and F. Pereira, The MPEG-4 Book. Englewood Cliffs, NJ: Prentice-Hall, 2002.
[20] F. Dufaux and T. Ebrahimi, “H.264/AVC Video Scrambling for Privacy Protection”, in Proc. IEEE International Conference on Image Processing, San Diego, CA, Oct. 2008.
[21] T. Wiegand, G.J. Sullivan, G. Bjøntegaard, and A. Luthra, “Overview of the H.264/AVC Video Coding Standard”, IEEE Trans. on Circuits and Systems for Video Technology, vol. 13, no. 7, pp. 560-576, July 2003.
[22] W. Zhang, S.S. Cheung, and M. Chen, “Hiding privacy information in video surveillance system”, in Proc. IEEE International Conference on Image Processing, Genoa, Italy, Sept. 2005.
[23] S.S. Cheung, J.K. Paruchuri, T.P. Nguyen, “Managing Privacy Data in Pervasive Camera Networks”, in Proc. IEEE International Conference on Image Processing, San Diego, CA, Oct. 2008.
[24] D. Choujaa and N. Dulay, “Towards Context-aware Face Anonymisation”, in Proc. 7th International Conference on Mobile and Ubiquitous Multimedia, Umea, Sweden, Dec. 2008.
[25] A. Massoudi, F. Lefebvre, C. De Vleeschouwer, B. Macq, and J.-J. Quisquater, “Overview on Selective Encryption of Image and Video: Challenges and Perspectives”, EURASIP Journal on Information Security, Volume 2008, Article ID 179290, doi:10.1155/2008/179290.
[26] Y. Mao and M. Wu, “A joint signal processing and cryptographic approach to multimedia encryption”, IEEE Trans. on Image Processing, vol. 15, no. 7, pp. 2061-2075, July 2006.
[27] M. Saini, P.K. Atrey, S. Mehrotra, S. Emmanuel and M. Kankanhalli, “Privacy Modeling for Video Data Publication”, in Proc. IEEE International Conference on Multimedia & Expo (ICME 2010), Singapore, July 2010.
[28] A. Said, “Measuring the Strength of Partial Encryption Schemes”, in Proc. IEEE International Conference on Image Processing, Genova, Italy, Sept. 2005
[29] M.A. Turk and A.P. Pentland, “Face Recognition Using Eigenfaces”, in Proc. of IEEE Conference on Computer Vision and Pattern Recognition, Maui, HI, June 1991.
[30] W. Zhao, R. Chellappa, and A. Krishnaswamy, “Discriminant analysis of principal components for face recognition”, in Wechsler, Philips, Bruce, Fogelman-Soulie, and Huang, editors, “Face Recognition: From Theory to Applications”, pp. 73–85, 1998.
[31] D. E. Knuth, “The Art of Computer Programming (volume 2)”. Reading, MA: Addison–Wesley, 1969.