IET Signal Processing

Research Article

IET Journals
The Institution of Engineering and Technology

Image privacy protection with secure JPEG transmorphing

ISSN 1751-9675
Received on 30th December 2016
Revised 13th July 2017
Accepted on 11th August 2017
doi: 10.1049/iet-spr.2016.0756
www.ietdl.org

Lin Yuan[1] ✉, *Touradj Ebrahimi*[1]

[1]*Multimedia Signal Processing Group, Electrical Engineering Department, EPFL, Station 11, Lausanne, Switzerland*
✉ *E-mail: lin.yuan@epfl.ch*

Abstract: Thanks to advancements in smart mobile devices and social media platforms, sharing photos and experiences has significantly bridged the authors' lives, allowing them to stay connected despite distance and other barriers. Most approaches to protect image visual privacy focus on encrypting or permuting image data, which generate unreadable image or highly distorted visual effect and therefore may not be in users best interest from both usage and perception perspectives. In this study, the authors propose secure JPEG transmorphing, a framework for protecting image visual privacy in a secure, reversible, and highly flexible and personalised manner. Secure JPEG transmorphing allows one to apply arbitrary regional visual manipulation on image regions of interests (ROIs), while secretly preserving the information about the original ROIs in application segments (APPn markers) of the visually obfuscated JPEG image. Objective and subjective experiments have been performed and results indicate that the proposed protection scheme provides near lossless image reconstruction, controllable level of file size expansion, good degree of privacy protection and especially better subjective pleasantness.

1 Introduction

The number of images shared from mobile devices has reached scales which were unimaginable only a decade ago: Every day over two billion images are posted to Online Social Networks (OSNs) or exchanged through instant messaging and cloud-based sharing services. However, advancements of photo sharing have also raised concerns for privacy, as photos potentially reveal great amount of sensitive information about people. OSN sites usually offer a limited degree of privacy protection and the most common solution is just conditional access. Researchers in the field of image processing and media security have proposed various approaches to enable photo privacy, most of which focus on encrypting or permuting the entire image data. From data security point of view, an encryption-based scheme can protect privacy in a highly secure and reversible manner. However, simply encrypting an entire image may significantly affect the usability of photo sharing and may not be in users interest from both usage and perception perspectives. In many cases, people seek simple and intuitive solutions to share their photos online to public while partially protecting specific regions in an image, e.g. creating an anonymous face with a cartoon smiley, blurring or inpainting sensitive areas. Yet, all those interesting manipulations cannot be reversed directly. Inspired by these facts, we attempt to explore novel solutions to protect image visual privacy in not only secure and reversible but also intuitive and pleasant ways.

In this paper, we present secure JPEG transmorphing, a novel framework for JPEG (Joint Photographic Experts Group) compression. Within secure JPEG transmorphing, most types of regional visual obfuscations can be applied, such as masking, blurring, pixelation, inpainting and warping. More importantly, the original image can be reconstructed with near lossless quality, even if the protected image has been manipulated. Objective experiments were conducted to evaluate the performance of the proposed method with respect to its reconstruction quality and storage overhead. Moreover, we also investigate the privacy protection capability and users pleasantness of different regional image obfuscations via a set of subjective experiments via online crowdsourcing.

The remaining of the paper is structured as follows. Section 2 presents related works. Section 3 describes in detail the working principle of the proposed protection framework. Then Sections 4 and 5 report the objective and subjective experiments, respectively. Finally, Section 6 concludes the paper.

2 Related work

Early-stage privacy protection mainly aims at enabling the visual privacy in video surveillance. The methods include image pixelation, masking, blurring [1, 2], scrambling [3], warping [4], and morphing [5]. With the rapid development of online social networks and photo sharing services, protecting privacy especially for online photos has raised great challenges. A number of research efforts focused on the design of access control mechanisms such that photos or data shared online can be only accessed by a selected group of people [6–10]. However, most existing access control systems are far from adequate, which could lead to severe information leakage when the OSN cannot apply the access control polices correctly, or if users fail to understand the complex privacy settings.

Another category of schemes to protect privacy in online images aim at encrypting or permuting the image data, before uploading and sharing them on OSN. The encryption or permutation can be performed in different domains, e.g. image pixels, bitstream, or discrete cosine transform (DCT) coefficients. Poller *et al.* [11] propose a robust image obfuscation, by permuting pixel blocks and modulating channel intensity. However, its security against different types of attacks is not clearly known and the permuted images could still reveal a certain degree of visual information. Ra *et al.* [12] propose P3, a photo privacy protection algorithm based on JPEG coding, which splits an image into a public part and a private part, with the public portion shared via OSN and private portion secretly stored in a storage server. However, P3 could disclose rich visual information in its public image when the threshold value is not small enough and its introduction of an additional cloud server complicates the file management system. Tierney *et al.* [13] propose a system named Cryptagram, which enables users to encrypt photos with traditional block ciphers and embed the encrypted bitstream into a JPEG file. However, this method creates a significant storage overhead due to the use of a cover image. Sun *et al.* [14] propose a privacy-aware regions of interest (ROI) image encryption scheme named Privacy-aware ROI Image Encryption based on logistical mapping and data hiding. The PRIE scheme utilises salient object detection to detect privacy-sensitive regions of an image. After encrypting pixels in

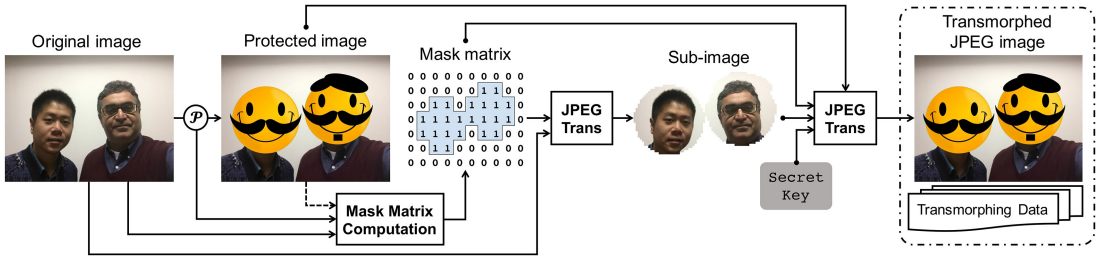

Fig. 1 *Protection procedure of secure JPEG transmorphing*

/∗ I_O: input original image; $\mathcal{P}$: obfuscation operation applied. ∗/
1: $I_P \leftarrow \texttt{obfuscate}(I_O, \mathcal{P}, \texttt{JPEG})$ # Obfuscated image I_P in JPEG
2: **if** $\mathcal{P}$ is known and well defined **then**
3: $\mathbf{M} \leftarrow \texttt{ComputeMaskMatrix}(I_O, \mathcal{P})$ # Compute $\mathbf{M}$ directly based on $\mathcal{P}$
4: **else**
5: $I_O^Y \leftarrow \texttt{RGBtoYUV}(I_O)[Y]$ # Y (luminance) channel of I_O
6: $I_P^Y \leftarrow \texttt{RGBtoYUV}(I_P)[Y]$ # Y (luminance) channel of I_P
7: $I_\Delta = |I_O^Y - I_P^Y|$
8: $I_\Delta^B = (I_\Delta > t)$
9: $\mathbf{M} \leftarrow \texttt{DownSample}(I_\Delta^B, 16)$ # Downsample I_Δ^B by factor of 16. If any pixel in an MCU block is 1, that block becomes an element 1 in $\mathbf{M}$. The mask matrix is round up to size of $(\lceil H/16 \rceil, \lceil W/16 \rceil)$ (W and H are width and height of image I_O).
10: **return** $\mathbf{M}$ # Generated mask matrix

Fig. 2 *Algorithm 1. GenerateMaskMatrix* ($I_O, \mathcal{P}$)

the private ROI using chaotic cryptography, the significant bits are embedded into the non-privacy region using data hiding. In addition, various approaches have been proposed to secure JPEG image based on DCT-domain encryption or scrambling. Yuan *et al.* [15, 16] propose a scrambling scheme that randomly changes the signs of DCT coefficients to enable image visual privacy. Recently, He *et al.* [17] proposed PUPPIES, a design and implementation of a partial image sharing technique, based on DCT-domain perturbation, which allows users to stipulate specific private image regions and correspondingly set different policies for each user. Other studies include [18, 19], both encrypting DCT coefficients with different schemes. Many of the solutions mentioned above support the protection of regional image information but some not, e.g. P3. However, all those solutions for protecting image visual privacy stay in the stage of encrypting or permuting image data, in either entire image or ROIs, which results in highly distorted visual effect with poor image quality. It may not be in users' best interest to see and use such methods to protect their own privacy, in particularly in the scenario of online social media. Yet, the impact of visual obfuscations on users' perception and usage preference has not been well studied nor understood.

3 Secure JPEG transmorphing: the framework

The working principle of secure JPEG transmorphing is to utilise JPEG application segments (APPn marker) to secretly preserve regional original image information, while encoding or transcoding the original image in a visually obfuscated form. The visual information can be protected by any type of regional image manipulation, e.g. masking, blurring, pixelation, inpainting and warping. The protected image, or called transmorphed image, of the same structure as standard JPEG, is backwards compatible with JPEG. With a dedicated JPEG transcoder or decoder that supports JPEG transmorphing, the original image can be recovered by replacing the obfuscated regions in the protected image with the corresponding original regions hidden in JPEG APPn markers.

3.1 Transmorphing protection

The protection procedure of secure JPEG transmorphing consists of three steps: (i) mask matrix generation, (ii) sub-image construction and (iii) transmorphing data insertion. Fig. 1 illustrates an example protection procedure for reader's easier understanding.

1: **while** Encoding/trancoding I_O to a new JPEG image I_{Sub} **do**
2: $(i_{\text{M}}, j_{\text{M}}) \leftarrow \texttt{IndexOfCurrentMCU}()$ # Index of MCU
3: **if** $\mathbf{M}(i_{\text{M}}, j_{\text{M}}) == 0$ **then** # Unprotected ROIs
4: $I_{\text{Sub}}.\texttt{MCUArray}(i_{\text{M}}, j_{\text{M}}) = 0$
5: **else** # Protected ROIs
6: $I_{\text{Sub}}.\texttt{MCUArray}(i_{\text{M}}, j_{\text{M}}) = I_O.\texttt{MCUArray}(i_{\text{M}}, j_{\text{M}})$
7: **return** I_{Sub} # Generated sub-image

Fig. 3 *Algorithm 2. ConstructSubImage* ($I_O, \boldsymbol{M}$)

3.1.1 Mask matrix generation: Firstly, user obfuscates certain ROIs of a given image using arbitrary regional manipulation, such as masking face regions with cartoon stickers as example images in Fig. 1. Upon the obfuscation applied, a binary-valued two-dimensional matrix is generated, which specifies the shape, size and position of protected ROIs. Each element of the matrix corresponds to a Minimum Coded Unit block of the upcoming encoded JPEG image, where elements 1 indicate protected blocks and 0 unprotected. This matrix is called *Mask Matrix*, noted as $\boldsymbol{M}$, which holds essential geometrical information about the image ROIs being protected. Depending on applications, the mask matrix can be generated either from geometrical information of user actions (e.g. coordinates of finger touch on mobile phone), or by comparing the original and obfuscated images. This step is presented in Algorithm 1 (see Fig. 2).

3.1.2 Sub-image construction: Secondly, a *sub-image* I_{Sub} is constructed by encoding or transcoding the original image to a new JPEG image, during which DCT coefficients corresponding to the ROIs defined by $\boldsymbol{M}$ are preserved while other coefficients outside the ROIs are set to zero. The sub-image is still a JPEG image with the same dimensions but smaller file size as the original image. It contains information about the original image ROIs to be protected. This procedure is presented in Algorithm 2 (see Fig. 3).

3.1.3 Transmorphing data insertion: Lastly, I_{Sub} is secured by a symmetric encryption scheme with a *secret key*, e.g. the Advanced Encryption Standard (AES) [20] or JPEG scrambling [15]. The bytestream of the encrypted sub-image C_{Sub}, the mask matrix $\boldsymbol{M}$, along with certain metadata, collectively called *Transmorphing data*, is then inserted into a set of application segments of the obfuscated JPEG image; in this respect, the obfuscated image serves as a 'cover image'. Here we encode the binary-valued mask matrix into a bitstream, with each bit representing an element of $\boldsymbol{M}$. The metadata contains the auxiliary information about the inserted sub-image and mask matrix, such as the data length and the encryption scheme and parameters applied. Since JPEG allows a maximum of 65,533 bytes [Each APP marker signals its marker length with two bytes (16 bits), resulting in $(2^{16} - 1) - 2 = 65,533$ bytes to record extra information.] allocated for each marker segment, the sub-image data very likely needs to be separately stored in several APPn segments. In our current implement, APP11 markers are used for signalling transmorphing data. This step is described in Algorithm 3 (see Fig. 4) and the syntax of a transmorphed image file is shown in Fig. 5.

/* I_P: obfuscated image in JPEG; $\mathcal{C}$: encryption scheme (with parameters); K: secret key. */

Insert metadata and mask matrix:

$C_{\text{Sub}} \leftarrow \texttt{Encrypt}(I_{\text{Sub}}, \mathcal{C}, K)$
$\text{BS}_{\mathbf{M}} \leftarrow \texttt{ByteStreamOf}(\mathbf{M})$
$\text{MD} \leftarrow [\texttt{SizeOf}(\text{BS}_{\mathbf{M}}),\ \texttt{SizeOf}(C_{\text{Sub}}),\ \mathcal{C}]$ # Metadata MD
$N_{(\text{MD},\mathbf{M})} \leftarrow \texttt{SizeOf}(\text{MD}) + \texttt{SizeOf}(\text{BS}_{\mathbf{M}})$ # Size of the first APP11 segment
I_P.`writeAPPnMarker`("0xFFEB") # A new APP11 marker
I_P.`writeMarkerLength`($N_{(\text{MD},\mathbf{M})}$) # Write marker length
I_P.`writeBytes`(MD)
I_P.`writeBytes`($\text{BS}_{\mathbf{M}}$)

Insert sub-image data:

$N_{\text{Sub}} \leftarrow \texttt{SizeOf}(C_{\text{Sub}})$
if $N_{\text{Sub}} <= 65533$ **then** # Write I_{Sub} data in one segment
    I_P.`writeAPPnMarker`("0xFFEB")
    I_P.`writeMarkerLength`(N_{Sub})
    I_P.`writeBytes`(C_{Sub})
else # Write I_{Sub} data in several segments
    $N_{\text{Marker}} = \lceil N_{\text{Sub}}/65533 \rceil$ # Number of segments
    for $i \in [1, \ldots, N_{\text{App}}]$ **do**
        I_P.`writeAPPnMarker`("0xFFEB")
        if $i \neq N_{\text{Marker}}$ **then**
            I_P.`writeMarkerLength`(65533)
            I_P.`writeBytes`(C_{Sub}.`byteArray`$[(i-1)*65533 : i*65533]$)
        else
            I_P.`writeMarkerLength`(N_{Sub} `mod` 65533)
            I_P.`writeBytes`(C_{Sub}.`byteArray`$[(i-1)*65533 : \text{end}]$)
return I_P # Final Transmorphed image I_P

Fig. 4 *Algorithm 3. InsertTransmorphingData* ($I_P, \boldsymbol{M}, I_{\text{Sub}}, \mathscr{C}, K$)

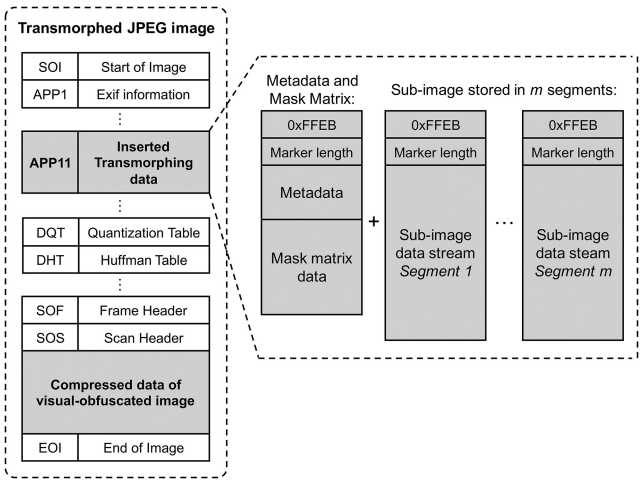

Fig. 5 *Syntax of a JPEG transmorphing protected image*

Overhead control (optional): Inserting additional information in protected image increases image file size and we therefore propose a simple mechanism to adjust such overhead by reducing the file size of the final transmorphed image based on DCT coefficients requantisation. To be more specific, we dequantise the DCT coefficients of the obfuscated ROIs in I_P and then requantise them with a new quality factor (noted as QF). The requantised DCT coefficients along with other original coefficients of unprotected ROIs are finally entropy-coded. DCT requantisation usually decreases image quality and reduces image file size, which has been identified and well explained in [21]. We just make use of such a 'side effect' of DCT requantisation to decrease the quality and data volume of the obfuscated regions presented in the 'cover image' I_P, further reducing the file size of transmorphed image.

3.2 Transmorphing reconstruction

The reconstruction procedure aims at recovering the original image from a transmorphed image, by reversing the above transmorphing

/* I_P: Transmorphed image; $\mathcal{T}$: image transformation applied on Transmorphed image; K: secret key to decrypt the sub-image. */

$(\mathbf{M}, C_{\text{Sub}}, \mathcal{C}) \leftarrow \texttt{ExtractTransmorphingData}(I_P)$
$\mathbf{M}_O \leftarrow \texttt{UpSample}(\mathbf{M})$ # Upsample mask matrix to the size of original image
$I_{\text{Sub}} \leftarrow \texttt{Decrypt}(C_{\text{Sub}}, \mathcal{C}, K)$
if $\mathcal{T}(\cdot)$ was applied, meaning $I_P = \mathcal{T}(I_P^O)$ **then**
    switch Type of the operation $\mathcal{T}(\cdot)$ **do**
        case Lossy Geometric Transformation: # Scaling, cropping, warping, etc.
            $I'_P \leftarrow \mathcal{T}^{-1}(I_P)$ # Reserve the transformation on protected image
            I'_P.`pixelArray`$[\mathbf{M}_O] = I_{\text{Sub}}$.`pixelArray`$[\mathbf{M}_O]$ # Pixels replacement
            $I_{\text{Rec}} \leftarrow \mathcal{T}(I'_P)$ # Apply $\mathcal{T}(\cdot)$ again to get the recovered image I_{Rec} of the same geometry as I_P
        case Lossy Compression: # E.g., JPEG compression
            I_P.`pixelArray`$[\mathbf{M}_O] = I_{\text{Sub}}$.`pixelArray`$[\mathbf{M}_O]$
            $I_{\text{Rec}} = I_P$
        case Lossless Rotation/Flipping via JPEG Transcoding:
            $\mathbf{M}' \leftarrow \mathcal{T}(\mathbf{M})$
            $I'_{\text{Sub}} \leftarrow \texttt{JPEGTranscoding}(I_{\text{Sub}}, \mathcal{T}(\cdot))$
            I_P.`MCUArray`$[\mathbf{M}'] = I'_{\text{Sub}}$.`MCUArray`$[\mathbf{M}']$
            $I_{\text{Rec}} = I_P$
else # If no transformation applied
    I_P.`MCUArray`$[\mathbf{M}] = I_{\text{Sub}}$.`MCUArray`$[\mathbf{M}]$
    $I_{\text{Rec}} = I_P$
return I_{Rec} # Reconstructed image

Fig. 6 *Algorithm 4. RecoverTransmorphing* ($I_P, \mathcal{T}(\cdot), K$)

protection operations: extracting C_{Sub} and $\boldsymbol{M}$ from inserted transmorphing data, decrypting C_{Sub} to get I_{Sub} and replacing the obfuscated ROIs in I_P with corresponding information in I_{Sub}. Since the inserted mask matrix and sub-image preserves the complete information about the original image corresponding to the protected ROIs, the protected image is robust to most image transformations. However, we assume that applied transformations do not remove the inserted data in JPEG header, and that the transformation is a known operation that can be re-applied. The replacement process can be done in either frequency (DCT coefficient) or spatial (pixel) domain depending on the transformation applied to the transmorphed image. The Algorithm 4 (see Fig. 6) presents the reconstruction procedure of secure JPEG transmorphing.

4 Objective performance evaluation

Image reconstructed from a reversible protection scheme is expected to be identical or highly similar to the original image. File size expansion due to inserted transmorphing data causes extra storage overhead, which should be as low as possible in practice. In this section, we evaluate the performance of the proposed secure JPEG transmorphing, in regard to the quality of reconstructed image and the storage overhead introduced. The evaluation was carried out in comparison with another JPEG-based algorithm, P3 [12].

4.1 Image dataset

The People In Photo Albums (PIPA) dataset [22] is used in our experiments. This is a dataset containing over 60,000 JPEG images of more than 2000 individuals collected from public Flickr photo albums. Each image of the dataset has ground truth head positions (rectangle) of several individuals annotated. We randomly selected a subset of 1500 images with the same size of 1204×768 pixels from the dataset and use this subset in our experiments [The subset used in this study is available at http://grebvm2.epfl.ch/lin/thesis/dataset/PIPA-subset-1500.zip]. In addition to the ground truth **Head** region, we consider two more ROIs as protection target: (i) the **Full-body** region, $3\times$ head width and $6\times$ head height, with head at the top centre of the full body and (ii) the **Upper-body** region, upper-half of the full body rectangle.

Table 1 Mean PSNR (dB) and SSIM of reconstructed images from JPEG transmorphing and P3 without and with transformations applied on protected images

Method	Without transformation		With transformation:			
			Sc.	Cr.	Co.	Ro.
secure JPEG transmorphing	PSNR	45.08	39.37	42.94	37.59	45.08
	SSIM	0.987	0.975	0.979	0.951	0.987
P3	PSNR	Inf.	37.46	43.27	35.53	inf.
	SSIM	1	0.968	0.993	0.925	1
baseline JPEG compression ($Q = 75$)					PSNR	36.32
					SSIM	0.9444

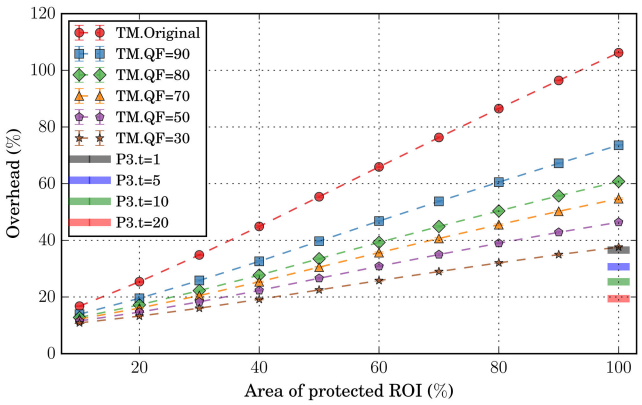

Fig. 7 *Storage overhead of secure JPEG transmorphing and P3*

4.2 Reconstruction quality

We applied secure JPEG transmorphing on an individual's three ROIs, respectively, in each image, and P3 on the entire image with a threshold of 5 (note that standard P3 only supports entire image protection). For secure JPEG transmorphing, we applied visual masking with a smiley sticker. Since spatial-domain masking operation involves JPEG decoding and re-encoding, which affects the quality of reconstructed images, we encoded the Transmorphed images with the maximal JPEG quality factor of 100 to minimise such impact. We applied four different transformations on each protected image and then executed the reconstruction process for each protected image without or with transformation applied. The four transformations are

- Scaling (Sc.): Subsample image by factor of 2 on both directions.
- Cropping (Cr.): Crop image to get centre region of size 512× 384.
- Compression (Co.): Recompress image in JPEG with QF 70.
- Rotation (Ro.): Rotate image by 90° in clockwise direction with a JPEG transcoder.

For P3, image transformations were applied on both public and secret portions. We use peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM) [23] to examine the quality of reconstructed image as compared to the original image. For scaled or cropped image, the two metrics were computed by comparing the reconstructed image with the original image manipulated by the same scaling or cropping operation.

The results of PSNR and SSIM for different reconstruction setups are shown in Table 1. For comparison, we also include the PSNR and SSIM of original images recompressed in JPEG with a quality factor of 75 as baseline. From the results, one observes the average PSNR and SSIM of reconstructed images from secure JPEG transmorphing are 45.08 dB and 0.987, which would be considered practically lossless in the signal processing community. With different transformations applied on the transmorphed images, reconstructed images still preserve significantly high quality, with PSNR and SSIM scores higher than that of the original images compressed in JPEG ($QF = 75$). As the rotation operation by JPEG transcoding is lossless, the quality measurements of corresponding reconstructed images are identical to that of images reconstructed without transformation applied. For P3, image reconstructed without transformation is lossless. With scaling and compression applied, the reconstruction quality is slightly worse than our method. This is because lossy image transformations may severely affect the DCT coefficients of P3 public and secret images, which decreases the precision of P3 reconstruction.

4.3 Storage overhead

To evaluate the storage overhead, we manually created ten mask matrices representing different ROIs of increasing area (from 10 to 100% of the entire image area) and applied secure JPEG transmorphing on each image with respect to each ROI, without and with DCT requantisation applied ($QF = 90$, 80, 70, 50 or 30). As JPEG decoding and re-encoding may affect image file size, we directly inserted the sub-image into the original JPEG image instead of creating the actual obfuscated image to diminish such impact. This is equivalent as if we assume the applied image obfuscation does not change the image file size. We also applied P3 [12] protection on each image, using four different threshold values (1, 5, 10 and 20). The definitions of storage overhead for secure JPEG transmorphing and P3 are given by the following equations, respectively:

$$\frac{S(I_{\mathrm{T}}) - S(I_{\mathrm{O}})}{S(I_{\mathrm{O}})} \text{ and } \frac{S\left(I_{\mathrm{P3}}^{\mathrm{Pub}}\right) + S\left(I_{\mathrm{P3}}^{\mathrm{Sec}}\right) - S(I_{\mathrm{O}})}{S(I_{\mathrm{O}})}, \tag{1}$$

where I_{O} denotes the original image, I_{T} is the transmorphed image, $I_{\mathrm{P3}}^{\mathrm{Pub}}/I_{\mathrm{P3}}^{\mathrm{Sec}}$ the public/secret part of P3 protected image and $S(\cdot)$ the operator to get image file size.

The storage overhead was computed for each protected image and results over all images (mean and 95% confidence interval) are presented in Fig. 7. From the result, a near linear relation between storage overhead and area of protected ROI is observed for different setups of secure JPEG transmorphing. Without DCT requantisation applied (TM.Original), the overhead of JPEG transmorphing is considerably higher, which is due to the sub-image data inserted. With DCT requantisation applied (TM.QF=n), such overhead is significantly reduced, e.g. by over 40% with $\mathrm{TM.QF} = 80$ compared to the case without overhead control. The overhead of P3 is in general lower than that of JPEG transmorphing when applied on the entire image region. Yet, the major purpose of using JPEG transmorphing is to protect regional image information, instead of obfuscating the entire image. For instance, by applying secure JPEG transmorphing on 40% of the entire image using QF of 80 or 70 in DCT requantisation, the overhead is only between 20 and 30%, which is acceptable in practise considering that JPEG image compressed with QF of 70 or 80 usually maintains fairly high quality.

5 Subjective evaluation of regional image obfuscation

A previous study [24] reveals that obfuscating just face region may not offer adequate protection to privacy against automatic person recognition carried out by deep neural networks employing visual cues from context information disclosed by other image regions

Fig. 8 *Reference and evaluation image sets of an example target*

Table 2 Visual privacy obfuscations applied in privacy evaluation

Name	Description
SCRB.H	high-level secure JPEG Scrambling [15]
SCRB.M	medium-level secure JPEG Scrambling [15]
P3.t = 5	regional P3 [12] protection with threshold 5
P3.t = 20	regional P3 [12] protection with threshold 20
blur	image blurring with radius of 8
pixelate	image pixelation with block size of 8
mask	visual masking with a smiley sticker on head or a grey rectangle on upper-/full-body region

unprotected. This challenges our method which is designed mainly for regional visual privacy protection. However, the degree to which different types of regional visual obfuscations can preserve users' privacy against real human 'attacking', namely recognition by humans, has not yet been well studied. Moreover, most visual obfuscations create unreadable or highly distorted image regions, the impact of which on users' pleasantness has not been identified. Therefore, we conducted a set of subjective experiments via crowdsourcing to investigate (i) the privacy protection capability and (ii) the pleasantness of regional image obfuscations.

5.1 Privacy protection capability

5.1.1 Methodology: We devised a novel subjective experiment where 'attackers' were put in a simplistic social networking scenario to evaluate the performance of different regional image obfuscations against person recognition from real human. The experiment was carried out by employing online subjects to act as attackers to recognise protected persons in images obfuscated by different visual protection methods in different setups. The lower the recognition rate is, the stronger the method is supposed to protect privacy. Amazon Mechanical Turk (AMT) [https://www.mturk.com/] was used as the crowdsourcing platform.

We selected six identities (adult male) from the PIPA dataset [22] as protection and recognition *targets* each having four images served as the *evaluation set*. We then applied seven obfuscations on three ROIs (defined in Section 4.1) of each target identity in each evaluation image, respectively. The seven obfuscations are described in Table 2. To increase the recognition difficulty, we selected another three identities in addition to the six identities such that subjects were required to identify protected target from a total of nine candidates. We designed three experiment setups to model different scenarios of person recognition 'attacking' in the context of social media. The three scenarios and corresponding setups are described in Table 3. The complete set of 24 evaluation images (in both original and protected versions) and all reference images of the 9 individuals (for both within-context and across-context scenarios) are available at http://grebvm2.epfl.ch/lin/privacy/privacy_dataset.zip. Fig. 8 shows the evaluation images and two sets of reference images of an example target.

Table 3 Person recognition scenarios and corresponding experiment setups

Scenario	Description	Experiment setup
Within-context person recognition	The attacker has rich prior knowledge about the protected target, e.g. the target's public photos in his online profile or the memory about the target if the attacker meets the target often.	Four unprotected images of each target are provided to subjects as reference (called *reference set*). The protected targets in evaluation set have similar or the same context (dressing, event, people nearby or environment) as their reference set.
Across-context person recognition	The attacker has somewhat prior knowledge about the protected target, but less straightforward, e.g. the target's public photos in a different photo album of different context information from the protected images.	Subjects are provided with a different set of reference images (four images/target) that have significantly different context information about the protected target (dressing, event, people nearby or environment) from the evaluation set.
Without-context person recognition	The attacker has limited prior knowledge about the protected target, e.g. a very vague memory about the target's facial appearance.	No any reference image is provided and subjects need to identify the protected target based on only merely profile head pictures of the nine candidates.

5.1.2 User study based on crowdsourcing: We conducted three subjective experiments on AMT with the above setups, respectively. Every experiment was divided into three sessions, each showing evaluation images with only one type of ROI protected. Every human intelligence task (HIT) of AMT presents six protected evaluation images and an extra image unprotected serving as 'honeypot' to help us remove sloppy subjects. Each image is associated with a question asking subjects to identify the target by selecting one from the nine candidates. The option 'I really don't know' can be selected if subject has no any clue about the protected person. For within-context and across-context setups, the 'honeypot' images were randomly selected from the reference images. For without-context setup, 'honeypot' images were only selected from the three extra identities apart from the six identities under evaluation. In such a way, the unprotected 'honeypot' images do not reveal any information about original evaluation images. For within-context and across-context setups, reference images of all nine candidates are presented in the beginning of each HIT and made always available during the experiment for subjects' review. Fig. 9 shows the screenshot of an image being evaluated in an HIT of AMT. All HITs were published on AMT with the following constraints satisfied:

- Six images evaluated in an HIT belong to six targets, respectively.
- The order in which the six targets and applied obfuscations appear is random.
- Each image or HIT is rated by at least seven different subjects.
- Every subject can take unlimited number of HITs within a session but cannot participate in more than one session.

Finally, after filtering out outliers who answered 'honeypot' questions incorrectly, we collected responses from a total of 241 subjects, each rating 39.2 images on average.

5.1.3 Results and analysis: Fig. 10 shows the average proportion of correct, incorrect, and 'I don't know' answers over all images in the evaluation set, with respect to different recognition scenarios and protection ROIs. First, one observes that the recognition accuracy (proportion of correct answers) of within-

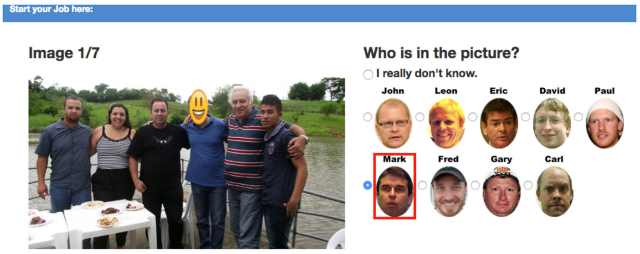

Fig. 9 *Screenshot of HIT presenting an image under evaluation*

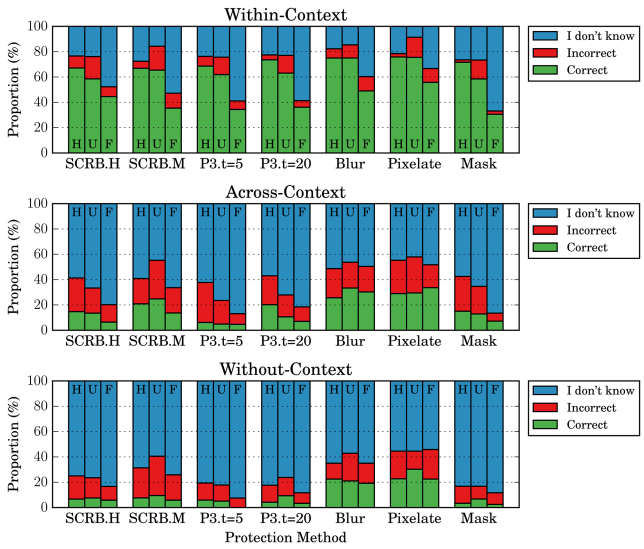

Fig. 10 *Proportion of 'I don't know', incorrect and correct answers across all images, with respect to different protection methods and regions. 'H', 'U' and 'F' annotated on each bar indicates head, upper-body and full-body, respectively*

context scenario remains high in most cases: For images with only head or upper-body protected, recognition accuracy against most obfuscations is above 60%; while for full-body protected images, the accuracy is significantly reduced, but still well above the level of random guess. The performances of all the seven obfuscations are comparable, with blur and pixelate showing slightly worse protection to privacy. We admit that in the case where direct context information about the protected person is available, regional visual obfuscation may not provide reliable privacy protection.

In the scenario of across-context recognition, overall recognition accuracies are greatly reduced. In turn, the proportions of incorrect and 'I don't know' answers are significantly raised. The accuracies corresponding to most methods such as JPEG scrambling, P3 and visual masking are lower than 20%. While the accuracies for blur and pixelate protections are obviously higher than the others, because the two methods still disclose certain low-resolution visual information about the original image region.

As for the without-context scenario, recognition accuracies are further reduced: the proportions of correct answers for protection methods including JPEG scrambling, P3 and visual masking are all below 10%. However, the accuracies for blur and pixelate are still higher than 20%. Among all the seven obfuscations, mask and P3.$t = 5$ provide the strongest protection against recognition, which is reasonable as masking operation completely hides the visual information behind the mask and a strong level of P3 protection is visually similar to a grey mask.

5.2 Pleasantness

5.2.1 Methodology: The second subjective experiment aims at understanding the pleasantness aspect of regional image visual obfuscations. First of all, we give the definition of the term 'pleasantness', which are considered in twofold: (i) *Perception Pleasantness* (users' perceived feeling when observing a photo obfuscated by a particular method) and (ii) *Usage pleasantness*

Fig. 11 *Example image protected by the ten visual obfuscations*
(a) SCRB, ***(b)*** P3, ***(c)*** Pixelate, ***(d)*** Blur, ***(e)*** Black, ***(f)*** Smiley, ***(g)*** TearsJoy, ***(h)*** SnapGhost, ***(i)*** Vendetta, ***(j)*** C-Stamp

Table 4 Ten visual privacy protection methods

Name	Description
SCRB	high-level JPEG Scrambling
P3	regional P3 [12] protection with a threshold of 20
Pixelate	image pixelation with block size of 20
Blur	image blurring with radius of 20
Black	visual masking in black colour
Smiley	visual masking with a 'smiley' Emoji
TearsJoy	visual masking with a 'tears of joy' Emoji
SnapGhost	visual masking with a Snapchat Ghost logo
Vendetta	visual masking with a cartoon Guy Fawkes mask originally from the film *V for Vendetta*
C-Stamp	visual masking with a grey stamp showing 'CONFIDENTIAL'

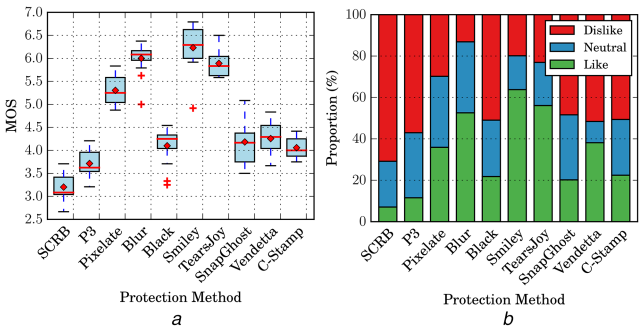

Fig. 12 *Results of subjective pleasantness measurements*
(a) Perception pleasantness, ***(b)*** Usage pleasantness

(users' preference to use a particular method to protect their own photo privacy). To measure perception pleasantness, we apply the Valence model in psychology with 9-Point SAM scales [25], where 1 stands for very unpleasant, 9 for very pleasant and the middle point 5 for neutral emotion. We use the three-level preference scales (i.e. 'dislike', 'neutral' and 'like') to model usage pleasantness. The subjective experiment was therefore conducted by gathering the two types of pleasantness responses on different visual protections as perceived by online subjects.

Ten visual protection obfuscations (listed in Table 4) were selected for comparison. An example image obfuscated by the ten methods is shown in Fig. 11. We selected 13 images from the PIPA dataset [22] and applied the 10 obfuscations on an individual's head region in each image, thus resulting in 130 different protected images [All images (original and protected) are available at http://grebvm2.epfl.ch/lin/privacy/pleasantness.zip]. For each protected image we asked 25 different subjects on AMT to vote on the perception and usage pleasantness. We removed the results from one subject who provided constant answers. Finally, 105 unique subjects participated in our experiment, each voting on 30.95 images on average. Screenshot of an example HIT on AMT is given at http://grebvm2.epfl.ch/lin/pleasantness_AMT.png.

5.2.2 Results and analysis: For each protected image, we computed its mean opinion score (MOS) of perception

pleasantness across ratings from different subjects. The MOSs of all 13 images corresponding to each protection method are shown as boxplot in Fig. 12*a*. From the result, one observes that pixelate, blur, smiley and TearsJoy Emoji provide significantly higher perception pleasantness scores than the others: the MOSs of pixelate, blur, smiley and TearsJoy are mostly above 5.0, indicating positive emotions; while other obfuscations reveal only negative emotions with average pleasantness scores below 5.0. Among the ten methods, JPEG scrambling (SCRB) and P3 provide the lowest perception pleasantness. We believe this is because image pixelation and blurring generate the most natural visual effects while the two Emoji stickers are the most enjoyable and amusing ones among all the methods. On the contrary, the two distortion-based approaches, JPEG scrambling and P3, can only result in the most unattractive visual effects. Interestingly, image stickers such as the SnapGhost and Vendetta still reveal relatively low pleasantness, though funny and interesting. The other two methods, Black masking and C-Stamp, generate similar level of perception pleasantness as SnapGhost and Vendetta.

As for usage pleasantness, Fig. 12*b* shows the overall proportions of votes for 'dislike', 'neutral' and 'like' of each protection method. Again, the two Emoji stickers smiley and TearsJoy obtained the most votes for 'like', above 60 and 55%, respectively. Pixelate and blur also received a large number of votes for 'like' (35 and 51%), and at the same time a large number of votes for 'neutral' (both around 30%). This time, the Vendetta mask got a considerable proportion of votes for 'like' and the smallest proportion of votes for 'neutral'. Compared to SnapGhost and C-Stamp, Vendetta received the same number of votes for 'dislike', indicating that the Vendetta mask is prone to being either liked or disliked by people. In addition, the other methods all received much less votes for 'like'. Among all the ten methods, the two distortion-based obfuscations, JPEG scrambling and P3, again received the largest number of votes for 'dislike', indicating the disadvantage of distortion-based obfuscations.

6 Conclusion

This paper presents secure JPEG transmorphing, a flexible framework for protecting image visual privacy in a secure, reversible and personalised manner. Secure JPEG transmorphing allows one to apply arbitrary regional visual manipulation on image ROIs, while secretly preserving the information about the original ROIs in application segments (APPn markers) of the visually obfuscated JPEG image. The protected image (or transmorphed image) has the same syntax as standard JPEG and is therefore backwards compatible with JPEG. With a dedicated JPEG transcoder or decoder that supports JPEG transmorphing, the original image can be recovered by replacing the obfuscated regions in the protected image with the corresponding original regions extracted from APPn markers. As arbitrary regional image manipulation can be applied, the proposed method provides a significant flexibility and usability such that users can choose their preferred ways to protect any sensitive image regions while still preserving its reversibility. This is the most distinctive characteristic of the proposed method compared to the others. In secure JPEG transmorphing, sensitive information represented in the sub-image is encrypted before being inserted. Therefore, the security of the proposed solution mostly relies on the encryption scheme applied. The security analysis is out of the scope of this paper.

With secure JPEG transmorphing, images of near lossless quality (compared to original image) can be reconstructed from protected image even if the latter has been manipulated by lossy transformations, e.g. scaling, cropping and compression. Inserting additional transmorphing data in JPEG image causes overhead to image file size but such overhead can be modulated by the proposed overhead control mechanism without affecting the reconstruction quality. Both facts are verified in our objective experiments.

Regional image obfuscations may not offer the perfect protection to privacy in certain scenarios as unprotected regions in image may still reveal private information, it is reliable enough in cases where unprotected image regions do not disclose straightforward context information about the protected target that matches other public information of that target. This is proven in a subjective experiment carried out via online crowdsourcing.

Last but not the least, we conducted another subjective experiment using crowdsourcing to investigate the pleasantness of different image obfuscations. Results indicate that distortion-based visual protections (e.g. JPEG scrambling and P3) may not provide the optimal pleasantness from both perception and usage perspectives. Instead, more intuitive, personalised and still reversible visual protection can be achieved using the proposed secure JPEG transmorphing.

7 Acknowledgments

This work was possible thanks to the research project LEADME (200020-149259) funded by Swiss National Foundation for Scientific Research and support from COST Action IC1206 DE-ID.

8 References

[1] Schiff, J., Meingast, M., Mulligan, D. K., *et al.*: 'Respectful cameras: detecting visual markers in real-time to address privacy concerns'. IROS, 2007, pp. 971–978

[2] Park, S., Trivedi, M.M.: 'A track-based human movement analysis and privacy protection system adaptive to environmental contexts'. AVSS, IEEE Computer Society, 2005, pp. 171–176

[3] Dufaux, F., Ebrahimi, T.: 'Video surveillance using JPEG 2000'. Proc. of the SPIE, 2004, vol. **5588**, pp. 268–275

[4] Korshunov, P., Ebrahimi, T.: 'Using warping for privacy protection in video surveillance'. 18th Int. Conf. on Digital Signal Processing (DSP), 2013

[5] Korshunov, P., Ebrahimi, T.: 'Using face morphing to protect privacy'. IEEE Int. Conf. on Advanced Video and Signal-Based Surveillance (AVSS), 2013

[6] Cutillo, L.A., Molva, R., Önen, M.: 'Privacy preserving picture sharing: enforcing usage control in distributed on-line social networks'. 5th ACM Workshop on Social Network Systems, SNS 2012, Bern, Switzerland, April 2012

[7] Klemperer, P., Liang, Y., Mazurek, M., *et al.*: 'Tag, you can see it!: using tags for access control in photo sharing'. Proc. of the SIGCHI Conf. Human Factors in Computing Systems, CHI '12, New York, NY, USA, 2012 pp. 377–386

[8] Mazurek, M.L., Liang, Y., Melicher, W., *et al.*: 'Toward strong, usable access control for shared distributed data'. Proc. of the 12th USENIX Conf. File and Storage Technologies (FAST 14), 2014, Santa Clara, CA, pp. 89–103

[9] Baden, R., Bender, A., Spring, N., *et al.*: 'Persona: an online social network with user-defined privacy', *SIGCOMM Comput. Commun. Rev.*, 2009, **39**, pp. 135–146

[10] Yuan, L., Theytaz, J.R., Ebrahimi, T.: 'Context-dependent privacy-aware photo sharing based on machine learning'. Proc. of 32nd Int. Conf. ICT Systems Security and Privacy Protection (IFIP SEC 2017), 2017

[11] Poller, A., Steinebach, M., Liu, H.: 'Robust image obfuscation for privacy protection in Web 2.0 applications', Proceedings of SPIE Volume 8303, Media Watermarking, Security, and Forensics 2012, 2012, 830304, doi: 10.1117/12.908587

[12] Ra, M.-R., Govindan, R., Ortega, A.: 'P3: toward privacy-preserving photo sharing'. Presented as part of the 10th USENIX Symp. Networked Systems Design and Implementation, Berkeley, CA, 2013, pp. 515–528

[13] Tierney, M., Spiro, I., Bregler, C., *et al.*: '*Cryptagram: photo privacy for online social media*' (Association for Computing Machinery, 2013), pp. 75–87

[14] Sun, J., Liao, X., Chen, X., *et al.*: 'Privacy-aware image encryption based on logistic map and data hiding', *Int. J. Bifurcation Chaos*, 2017, **27**, (05), p. 1750073

[15] Yuan, L., Korshunov, P., Ebrahimi, T.: 'Secure JPEG scrambling enabling privacy in photo sharing'. Workshop on De-Identification for Privacy Protection in Multimedia, 2015

[16] Yuan, L., Korshunov, P., Ebrahimi, T.: 'Privacy-preserving photo sharing based on a secure JPEG'. 2015 IEEE Conf. Computer Communications Workshops (INFOCOM WKSHPS), April 2015, pp. 185–190

[17] He, J., Liu, B., Kong, D., *et al.*: 'PUPPIES: transformation-supported personalized privacy preserving partial image sharing'. 2016 46th Annual IEEE/IFIP Int. Conf. Dependable Systems and Networks (DSN), June 2016, pp. 359–370

[18] He, K., Bidan, C., Le Guelvouit, G.: 'Robust and secure image encryption schemes during JPEG compression process'. 2016 IS&T Int. Symp. Electronic Imaging (EI 2016), CA, United States, February 2016

[19] Sun, W., Zhou, J., Lyu, R., *et al.*: 'Processing-aware privacy-preserving photo sharing over online social networks'. Proc. of the 2016 ACM on Multimedia Conf., MM '16, NY, USA, 2016, pp. 581–585

[20] Daemen, J., Rijmen, V.: '*The design of Rijndael*' (Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2002)

[21] Moon, J.W., Lee, J.S., Cho, N.I.: 'A requantization algorithm for the transcoding of JPEG images', *Signal Process., Image Commun.*, 2006, **21**, (1), pp. 13–21

[22] Zhang, N., Paluri, M., Taigman, Y., *et al.*: 'Beyond frontal faces: improving person recognition using multiple cues'. 2015 IEEE Conf. Computer Vision and Pattern Recognition (CVPR), June 2015, pp. 4804–4813
[23] Wang, Z., Bovik, A.C., Sheikh, H.R., *et al.*: 'Image quality assessment: from error visibility to structural similarity', *IEEE Trans. Image Process.*, 2004, **13**, pp. 600–612
[24] Oh, S.J., Benenson, R., Fritz, M., *et al.*: '*Faceless person recognition: privacy implications in social Media*' (Springer International Publishing, Cham, 2016), pp. 19–35
[25] Bradley, M.M., Lang, P.J.: 'Measuring emotion: the self-assessment manikin and the semantic differential', *J. Behav. Therapy Exper. Psychiatry*, 1994, **25**, (1), pp. 49–59