FACIAL EXPRESSION PRESERVING PRIVACY PROTECTION USING IMAGE MELDING

Yuta Nakashima[1], *Tatsuya Koyama*[2], *Naokazu Yokoya*[1], *Noboru Babaguchi*[2]

[1]Graduate School of Information Science
Nara Institute of Science and Technology
8916-5 Takayamacho, Ikoma, Nara, Japan

[2]Graduate School of Engineering
Osaka University
2-1 Yamadaoka, Suita, Osaka, Japan

ABSTRACT

An enormous number of images are currently shared through social networking services such as Facebook. These images usually contain appearance of people and may violate the people's privacy if they are published without permission from each person. To remedy this privacy concern, visual privacy protection, such as blurring, is applied to facial regions of people without permission. However, in addition to image quality degradation, this may spoil the context of the image: If some people are filtered while the others are not, missing facial expression makes comprehension of the image difficult. This paper proposes an image melding-based method that modifies facial regions in a visually unintrusive way with preserving facial expression. Our experimental results demonstrated that the proposed method can retain facial expression while protecting privacy.

Index Terms— Visual privacy protection, social networking services, facial expression, image context

1. INTRODUCTION

Currently, an enormous number of images are uploaded to social networking services such as Facebook and Twitter, which are then browsed and viewed by a vast amount of the Internet users. A critical concern about these social images is that they usually contain people's appearances, which are highly privacy sensitive. Thus one cannot share such images as is through SNSs without permission from each person. Otherwise, the uploaded images may violate the people's privacy.

Generally, to prevent from violating other people's privacy, one may apply an image processing technique, e.g., blocking out and blurring to remove the sensitive information such as facial regions. For example, Google Street View blurs all facial regions in its images, which can be done fully automatically [1, 2]. Video surveillance is an actively studied application of visual privacy protection [3, 4, 5], and some systems selectively filter sensitive regions [6, 7].

For social images/video, selective filtering of sensitive region as shown in Fig. 1(a) and (b) is inevitable; one may obtain permission for capturing and publishing an image from some people (and thus, corresponding image regions are disclosable) but not from the others. Basically, this selection have been done manually since who are disclosable and who are not can be highly dependent, although recent research efforts realized an automatic technique, focusing on the videographer's intentions [8].

(a) (b)

Fig. 1. Example privacy protected images blocking out (a) and blurring (b), where the left and right persons are non-disclosable and disclosable, respectively.

Unfortunately, such privacy protection suffers from visual artifacts and loss of facial expression. In Fig. 1(a), blocking out completely removes the facial detail of the non-disclosable person and causes severe visual artifacts. Figure 1(b) also loses the facial expression almost completely, and artifacts due to blurring is still significant. Among these problems, the loss of facial expression can be critical for social images because facial expression serves as an essential cue to comprehend the atmosphere or situation of the people.

Some methods can remedy this problem. Peng et al. [9] and Tanaka et al. [10] proposed to use avatars instead of presenting privacy protected faces. They track facial features or body parts so that avatars can be faithful to the facial expression and body configuration of the original person. A drawback of such methods is artificiality of obtained images as they superimpose computer graphics-based avatars on images. Another possible remedy is use of morphing, which mixes two face images together using geometric transformation. Unfortunately, the existing method [11] per se is not designed to retain facial expression, and the mixed face image

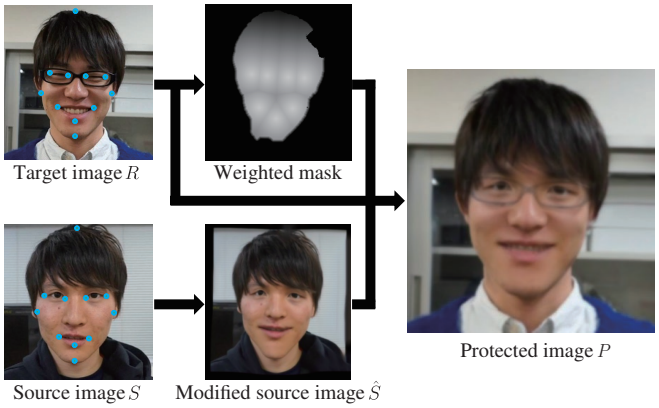

Fig. 2. Overview of our proposed method.

lies somewhere in-between the two faces. In addition, such a morphing-based method requires many corresponding points. This is cumbersome without help of automatic facial feature detection [12], which is still challenging under uncontrolled environments.

This paper proposes a method for image privacy protection, aiming at visually unintrusive privacy protection for facial images with preserving facial expression. The proposed method shares the same basic idea as morphing-based privacy protection because it modifies a person's face (a target image) by mixing other person's face (a source image) provided by the user. The proposed method is built upon image melding [13], and we modify the algorithm to retain facial expression of non-disclosable people. For this, we leverage a smaller number of corresponding points in target and source images specified by the user as a prior for facial feature positions. The primal contribution of this work is thus this modified algorithm. We experimentally demonstrate the performance of the proposed method, comparing with two baseline methods.

2. IMAGE MELDING-BASED PRIVACY PROTECTION

Figure 2 shows an overview of the proposed method. A user first inputs target image R and source image S, together with corresponding points on them, where the target image contains a facial region to be protected (multiple face regions can be handled by repeatedly applying the proposed method to each face region). The source image, chosen by the user, contains another person's face that is in similar orientation to the face in the target image. Using our image melding-based algorithm, the proposed method modifies the source image according to the corresponding points, so that ones on the source image coincide with those on the target image. It also generates a weighted mask based on the target images and the corresponding points on it. The proposed method synthesizes a privacy protected image by mixing the target image and the modified source image using the mask.

The proposed method assumes the following application scenario. When a user wants to take a image of people, the user asks for permission to share the image through a SNS. Some people may provide the user with the permission and the others may not. The user even has no contact with some people like passers-by. Therefore, she/he removes the sensitive information regarding these people from the image. For this, the user picks some appropriate facial images up from her/his dataset, applying our proposed method to the image and uploading it to the SNS.

In this scenario, we presume that the user has a dataset that contains facial images. A preliminary built set of de-identified faces [14] is potentially applicable to the proposed method without any privacy concern. This work, however, deals with a severer case where a manually built facial image dataset is used and both target and source images are privacy sensitive. The capability of privacy protection in this method is objectively and subjectively evaluated in Section 3.

2.1. Source image modification

For preserving facial expression, the proposed method modifies the source image according to the corresponding points on it using our image melding-based algorithm. Because the configuration of facial features, such as the eyes and mouth, is considered to be crucial for facial expression, the user are asked to specify 11 corresponding points as indicated by blue points in the leftmost images of Fig. 2, which roughly correspond to the edges or corners of the facial features and the face contour. For non-frontal faces, the user only specifies visible points.

Image melding [13] is a technique that can reproduce an image using patches from another image. For each patch in the former image, it searches the latter image for the most similar patch, and generates a smooth image that preserves discontinuity in the original images using the patch. Instead of using two images, our modified algorithm reproduces the source image using the patches from the same source image, with shifting the patches around the corresponding points on the source image to around the positions of the target image's corresponding points. By doing this, the positions of facial features on the source image match the target image's based only on a smaller number of corresponding points than the morphing-based method [11]. The detail of source image modification is as follows.

The proposed method first applies an affine transformation to the source image so that the corresponding points in it roughly match to those on the target image. This transformation adjusts the scale, orientation, and position of facial features. Although image melding allows a local geometric transformation during patch search, we disable it in this method because it can cause inconsistent sizes of facial features (e.g., different sizes of left and right eyes). Let $\mathbf{x}_k$ and $\mathbf{y}_k$ denote the k-th corresponding points on the source image

S and the target image R, respectively. The affine transformation (the linear component $\mathbf{A}$ and the translation component $\mathbf{b}$) is obtained by

$$\mathbf{A}^*, \mathbf{b}^* = \arg\min_{\mathbf{A},\mathbf{b}} \sum_k \|\mathbf{y}_k - (\mathbf{A}\mathbf{x}_k + \mathbf{b})\|^2. \quad (1)$$

We globally transforms S by this transformation. The transformed image is denoted by S', and $\mathbf{x}'_k = \mathbf{A}^*\mathbf{x}_k + \mathbf{b}^*$.

The proposed method then modifies S' to generate modified source image $\hat{S}$ by precisely adjusting the positions of the corresponding points. The algorithm is based on Algorithm 1 in [13], but we introduce an additional step that replaces pixels around $\mathbf{y}_k$ in intermediate image T and gradient image ∇T with pixels around $\mathbf{x}'_k$ in S' after the "ReconstructImage" step. Let $\alpha_{\mathbf{y}_k}(\mathbf{x})$ be a weight that gives 1 if $\mathbf{x} = \mathbf{y}_k$, linearly decreases as $\|\mathbf{x}-\mathbf{y}_k\|$ increases, and gives 0 if $\|\mathbf{x}-\mathbf{y}_k\| > \theta$, where θ is a parameter. Using this weight, we replace the pixels around each $\mathbf{y}_k$ by

$$T'(\mathbf{x}) = \alpha_{\mathbf{y}_k}(\mathbf{x})T(\mathbf{x}) + \{1 - \alpha_{\mathbf{y}_k}(\mathbf{x})\}S(\mathbf{x}) \quad (2)$$

$$\nabla T'(\mathbf{x}) = \alpha_{\mathbf{y}_k}(\mathbf{x})\nabla T(\mathbf{x}) + \{1 - \alpha_{\mathbf{y}_k}(\mathbf{x})\}\nabla S(\mathbf{x}), \quad (3)$$

and use T' and $\nabla T'$ instead of T and ∇T in the following process. Although this replacement may lead to discontinuity in T' and $\nabla T'$, iterative updates reduce it.

2.2. Mixing

The proposed method mixes the target image R and the modified source image $\hat{S}$ to obtain protected target image P, which is expected to reduce privacy disclosure of the person in the source image. Our source image modification re-arranges the facial features so that their positions coincide with those in the target image; however, their details, such as shape and sizes, are different. In addition, non-facial regions in the target image R should not change. We thus generate a weighted mask based on the corresponding points.

To keep non-facial regions unchanged, the proposed method extracts the facial region in R using a graph cut-based method. This method uses the corresponding points instead of interactive user input in GrabCut [15]. We then assign a weight to each pixel in the extracted region. In order to reduce duplicated facial features due to mixing, we design the weight so that pixels around each corresponding point mainly come from $\hat{S}$. More specifically, weight $\beta(\mathbf{x})$ for the pixel at $\mathbf{x}$ is given by

$$\beta(\mathbf{x}) = \sum_{\mathbf{y}_k \in Y} f_{\mathbf{y}_k}(\mathbf{x}), \quad (4)$$

where Y is a subset of the corresponding point, i.e., we exclude the corresponding points on the contour of facial region from Y because they do not causes facial feature duplication. $f_{\mathbf{y}_k}(\mathbf{x})$ is defined using predefined parameter ρ and the Euclidean distance between the $\mathbf{y}$ and $\mathbf{x}_k$ as

$$f_{\mathbf{y}_k}(\mathbf{x}) = \begin{cases} 1 - \|\mathbf{x} - \mathbf{y}_k\|/\rho & \text{for } \|\mathbf{x} - \mathbf{y}_k\| < \rho \\ 0 & \text{otherwise} \end{cases}. \quad (5)$$

Table 1. Parameter values.

	(i)	(ii)	(iii)
κ for blurring	11	15	21
α for morphing	0.4	0.5	0.6
ρ for proposed method	96	144	192

The pixel value $P(\mathbf{x})$ in protected image P is given by

$$P(\mathbf{x}) = \beta(\mathbf{x})R(\mathbf{x}) + \{1 - \beta(\mathbf{x})\}S(\mathbf{x}). \quad (6)$$

3. EXPERIMENTAL RESULTS

We objectively and subjectively verified the proposed method's capability for privacy protection using a face recognition technique and questionnaire. Our subjective survey also show whether it retains facial expression. We also subjectively surveyed if protected images are visually intrusive. Our dataset contains 8 subjects' frontal faces in happiness, disgust, and neutral expressions as well as other 8 subjects' profile face in the same facial expressions.

For comparison, we employed blurring and morphing. The Gaussian kernel with size κ was used for blurring. For morphing, since the work [11] did not detail the positions of 21 corresponding points, we alternatively used 27 corresponding points in [16] for frontal faces. Since no literature has reported corresponding points for profile face, we used 21 points on the contour of a face and around facial features. After Delauney triangulation, our implementation of morphing locally transformed the source image so that the points on it coincided their corresponding points on the target image. The transformed image was mixed with the target image with a constant weight α. Compared with [11], which transforms both source and target images to match the corresponding points, this implementation is expected to preserve facial expression. The parameter value θ to determine the weight for pixel replacement in Section 2.1 is set to 5% of the width of source images (face regions). We used three parameter values for each method, which are summarized in Table 1. For all these parameters, a larger value changes target images more, which can provide stronger privacy protection.

3.1. Capability for privacy protection

For objectively evaluating the capability of the proposed method for privacy protection, we used a face recognition technique that uses local binary pattern histograms (LBPHs) [17]. For training the recognizer, we collected 24 facial images for each subjects besides the dataset for evaluation. These 24 facial images consist of two frontal and profile faces in happiness, disgust, and neutral expressions taken under indoor and outdoor environments. We applied the privacy protection methods to the facial images of each subject in happiness and disgust expressions in the evaluation dataset (32 facial images in total), which were fed to the recognizer. As

Table 2. Recognition rates. The numbers in parentheses are correct recognitions among 32 trials.

Parameter	(i)	(ii)	(iii)
Blurring	0.53 (17)	0.56 (18)	0.38 (12)
Morphing (Target)	0.53 (17)	0.41 (13)	0.38 (12)
Morphing (Source)	0.09 (3)	0.19 (6)	0.31 (10)
Proposed (Target)	0.44 (14)	0.31 (10)	0.22 (7)
Proposed (Source)	0.16 (5)	0.19 (6)	0.25 (8)

Fig. 3. Example images used in our subjective evaluation of the capability of privacy protection.

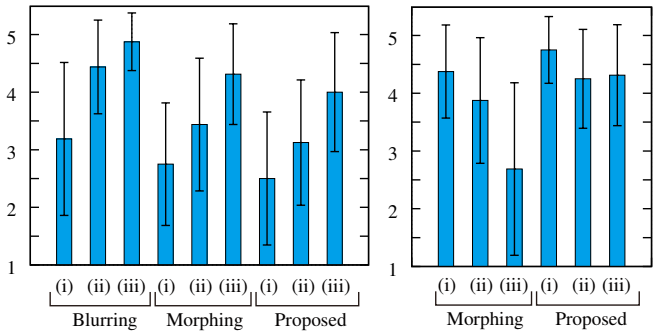

Fig. 4. Questionnaire results for the capability of privacy protection in (left) Scenario (A) and (right) Scenario (B). Results for blurring is not presented for the target images because it does not use target images.

morphing and the proposed method use face images for privacy protection, we calculated recognition rate (the number of correct recognitions divided by the number of recognition trials) for both source and target images. The image size was 160×160 pixels.

Table 2 summarizes the results. The recognition rate was 0.88 (28/32) for original face images. These results indicate that the capability of blurring for privacy protection is limited

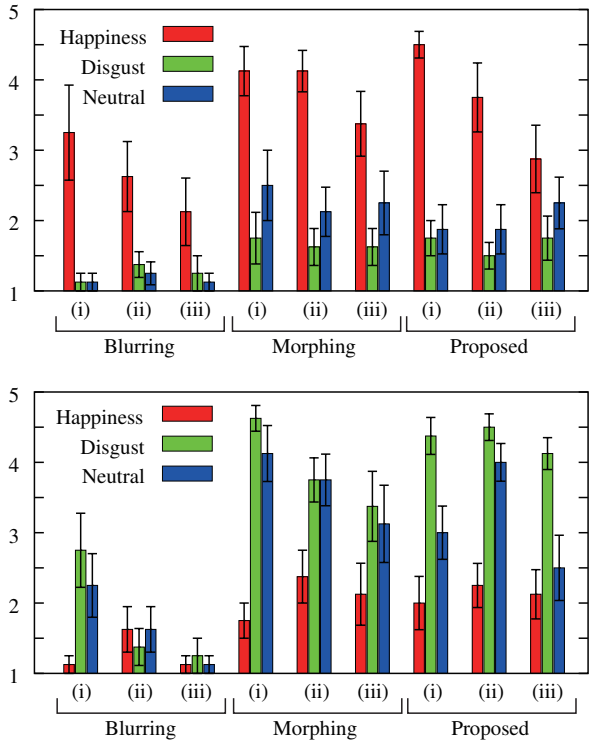

Fig. 5. Average scores and their standard deviation of questionnaire results on capability for preserving facial expression. Original images with happiness (top) and disgust (bottom) expressions.

even when $\kappa = 21$. The rates for target and source images by morphing with (iii) was the closest (around 0.35) among three parameter values, and those by the proposed method with (iii) was also closest (around 0.24). We consider the difference in these rates implies that the privacy protection capability of the proposed method is superior to morphing, at least for the LBPH-based recognizer.

Considering the subjectivity of privacy, we also used questionnaires for evaluating the capability for privacy protection. Since morphing and the proposed method require source images, our questionnaire consisted of two scenarios.

(A) To evaluate the capability when the subject is the target to be protected, the proposed method and the two baselines were applied to frontal face images with happiness and disgust expressions of each subject. The source image for morphing and the proposed method was randomly selected one with frontal face and neutral expression from our dataset.

(B) To evaluate the capability when the subject's frontal face images with happiness and disgust are used as a source image, morphing and the proposed method were applied to randomly selected target image with frontal face and neutral expression. Blurring does not apply to this scenario because it does not use source images.

Fig. 6. From left to right: examples of original images, protected images by blurring, morphing, and our proposed method. Right-most column shows source images used for morphing and our proposed method. From top to bottom: IMG1, IMG2, and IMG3. Red and blue rectangles indicate disclosable and non-disclosable people, respectively.

Each subjects reviewed facial images by the proposed method and the baselines with the parameters listed in Table 2, were asked if they feel their privacy is protected in the images, and assigned to each image a score ranging from 1 (completely disclosed) to 5 (completely protected). Some example face images used in the evaluation are shown in Fig. 3.

Figures 4(left) and (right) are the results for Scenarios (A) and (B), respectively. The results demonstrate that the capability of the proposed method with parameter (iii) is slightly worse than that for blurring with (ii) and is almost the same as morphing with (iii). Meanwhile, our subjects mostly felt that their privacy is protected with the proposed method when their facial image is used as a source image. One of the reason is that the proposed method retains the shape of faces (the facial contours) while morphing gradually changes it as well as the facial features.

3.2. Capability for preserving facial expression

We subjectively surveyed the capability of the proposed methods for preserving facial expression by questionnaire. The participants were exposed to their own images in happiness and disgust protected by blurring, morphing, and the proposed method, and judged whether the facial expression in the protected image was the same as that in the different face image in happiness, disgust, or neutral facial expressions without privacy protection. A score ranging from 1 to 5 was assigned to each protected facial image, where 1 stands for the protected facial image is completely different from the original image and 5 stands for completely the same facial expression as the original image. The source images for morphing and the proposed method were neutral face images that were randomly selected from frontal faces of the other 7 subjects.

Figure 5 shows the averages and standard deviations of the obtained scores. Morphing and the proposed method preserved facial expression compared to blurring. When the participants watched the protected image in happiness expression, the average score of the proposed method for happiness expression decreased more rapidly than morphing. One reason is that the proposed method can duplicate facial features because it automatically finds similar patches and uses them to fill the regions far from the corresponding points, which may make recognition of facial expression difficult. For original faces in disgust expression, the participants gave relatively high scores to neutral expression, because disgust expression is sometimes hard to differentiate from neutral expression.

3.3. Subjective evaluation of visual intrusiveness

We subjectively evaluated the visual intrusiveness of the protected image. We used three target images (IMG1, IMG2, and IMG3) shown in Fig. 6. In the original images, red rectangles and blue rectangles indicate disclosable and non-disclosable people, respectively. We asked 10 participants if they think each protected image was visually unintrusive, and they assigned a five point scale score to each image, where score 1 stands for completely disagree and 5 stands for completely agree. The results are shown in Fig. 7. The average scores for blurring, morphing, and the proposed method were 2.0, 3.9, and 4.2, respectively. The proposed method outperformed blurring and stably gave high average scores for all images.

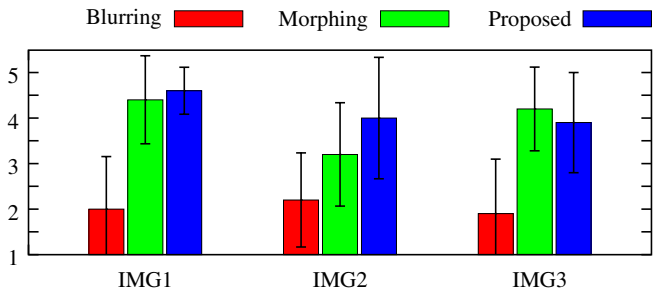

Fig. 7. Questionnaire results on visual intrusiveness.

4. CONCLUSION

In this paper, we proposed a method for privacy protection, which is designed to be visually unintrusive and to preserve the facial expression by using image melding [13]. We confirmed the proposed method's capability for privacy protection. We also subjectively evaluated whether the proposed method retained the facial expression and whether privacy protected images were visually unintrusive. As the results, the proposed method outperformed blurring, which is extensively used for conventional privacy protection. In addition, its capability for preserving facial expression is comparable to or better than morphing, while the number of corresponding points is smaller than morphing. Our future work includes implementing an automatic algorithm to detect facial features and to obtain corresponding points. Although we consider that the proposed method works with only acceptable amount of manual corresponding point specification, more corresponding points may improve the visual quality. Currently, we focus only on facial images, but it is also interesting to apply our idea to entire body regions, as well as other potential identifiers discussed in [18, 19]. This work was partly supported by JSPS Grants-in-Aid for Scientific Research and for Young Scientists.

5. REFERENCES

[1] A. Flores and S. Belongie, "Removing pedestrians from google street view images," in *Proc. IEEE Int. Workshop on Mobile Vision*, 2010, pp. 53–58.

[2] A. Frome, G. Cheung, A. Abdulkader, M. Zennaro, B. Wu, A. Bissacco, H. Adam, H. Neven, and L. Vincent, "Large-scale privacy protection in google street view," in *Proc. IEEE Int. Conf. Computer Vision*, 2009, pp. 2373–2380.

[3] J. Wickramasuriya, M. Datt, S. Mehrotra, and N. Venkatasubramanian, "Privacy protecting data collection in media spaces," in *ACM Int. Conf. Multimedia*, 2004, pp. 48–55.

[4] A. Senior, S. Pankanti, A. Hampapur, L. Brown, Y.-L. Tian, and Ahmet Ekin, "Enabling video privacy through computer vision," *IEEE Security & Privacy*, vol. 3, no. 3, pp. 50–57, 2005.

[5] F. Dufaux and T. Ebrahimi, "H.264/AVC video scrambling for privacy protection," in *Proc. IEEE Int. Conf. Image Processing*, 2008, pp. 1688–1691.

[6] X. Yu, K. Chinomi, T. Koshimizu, N. Nitta, Y. Ito, and N. Babaguchi, "Privacy protecting visual processing for secure video surveillance," in *Proc. IEEE Int. Conf. Image Processing*, 2008, pp. 1672–1675.

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

[8] Y. Nakashima, N. Babaguchi, and J. Fan, "Automatically protecting privacy in consumer generated videos using intended human object detector," in *Proc. ACM Int. Conf. Multimedia*, 2010, pp. 1135–1138.

[9] J. Peng, N. Babaguchi, H. Luo, Y. Gao, and J. Fan, "Constructing distributed hippocratic video databases for privacy-preserving online patient training and counseling," *IEEE Trans. Information Technology in Biomedicine*, vol. 14, no. 4, pp. 1014–1026, 2010.

[10] K. Tanaka, S. Onoue, H. Nakanishi, and H. Ishiguro, "Motion is enough: how real-time avatars improve distant communication," in *Proc. Int. Conf. Collaboration Technologies and Systems*, 2013, pp. 465–472.

[11] P. Korshunov and T. Ebrahimi, "Using face morphing to protect privacy," in *Proc. IEEE Int. Conf. Advanced Video and Signal Based Surveillance*, 2013, pp. 208–213.

[12] M. Dantone, J. Gall, G. Fanelli, and L. Gool, "Real-time facial feature detection using conditional regression forests," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2012, pp. 2578–2585.

[13] S. Darabi, E. Shechtman, C. Barnes, D. B. Goldman, and P. Sen, "Image melding: combining inconsistent images using patch-based synthesis," *ACM Trans. Graphics*, vol. 31, no. 4, pp. 82:1–82:10, 2012.

[14] E. M. Newton, L. Sweeney, and B. Malin, "Preserving privacy by de-identifying face images," *IEEE Trans. Knowledge and Data Engineering*, vol. 17, no. 2, pp. 232–243, 2005.

[15] C. Rother, V. Kolmogorov, and A. Blake, ""GrabCut": interactive foreground extraction using iterated graph cuts," *ACM Trans. Graphics (Proc. ACM SIGGRAPH)*, vol. 23, no. 3, pp. 309–314, 2004.

[16] R. Szeliski, *Computer Vision: Alrorithms and Applications*, chapter 3: Image Processing, Springer, 2011.

[17] T. Ahonen, A. Hadid, and M. Pietikainen, "Face description with local binary patterns: application to face recognition," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 28, no. 12, pp. 2037–2041, 2006.

[18] Mukesh Saini, Pradeep K. Atrey, Sharad Mehrotra, and Mohan Kankanhalli, "W^3-privacy: understanding what, when, and where inference channels in multi-camera surveillance video," *Multimedia Tools and Applications*, vol. 68, no. 1, pp. 135–158, 2014.

[19] Mukesh Saini, Pradeep K. Atrey, Sharad Mehrotra, and Mohan Kankanhalli, "Privacy aware publication of surveillance video," *Int. J. Trust Management in Computing and Communications*, vol. 1, no. 1, pp. 23–51, 2013.