Image Obfuscation with Quantifiable Privacy

Liyue Fan
University at Albany - SUNY
liyuefan@albany.edu

Abstract

There is an increased concern about image privacy nowadays, due to the popularity of camera-equipped devices. However, widely adopted image obfuscation solutions do not offer formal privacy guarantees. Recent approaches with k-anonymity or standard ϵ-differential privacy may not be suitable for practical applications. In this abstract, we describe a novel image obfuscation method based on metric privacy, a rigorous privacy notion generalized from differential privacy. Compared to existing approach with standard differential privacy, our solution achieves a balanced trade-off between privacy and utility by providing indistinguishability based on image visual similarity. Furthermore, our obfuscation solution is lightweight and does not require training, demonstrating feasibility for image sanitization on personal devices.

1. Introduction

Image obfuscation has been widely applied to obscure regions-of-interest (ROIs) in an image, such as faces and texts, when sharing image data with untrusted parties. However, due to rapid development in machine learning, standard obfuscation methods, such as pixelization and blurring, are no longer sufficient in privacy protection. For instance, McPherson et al. [6] showed that up to 96% of obfuscated faces can be re-identified by convet-based models.

More sophisticated obfuscation methods have been proposed to enhance privacy and utility. For instance, GANs (generative adversarial nets) has been adopted for image obfuscation, e.g., by *inpainting* the head region with an unknown, natural-looking identity as in Sun et al. [9], and by modifying the identity while preserving action detection as in Ren et al. [8]. Such approaches may heavily rely on training data, yet do not provide formal privacy guarantees. Newton et al. [7] proposed the k-Same algorithm for face de-identification. While achieving k-anonymity, this algorithm may be challenging to apply, as it requires a trusted server or collaboration among users. In our recent work [3], we proposed a pixelization method which satisfies $\epsilon-$differential privacy [2]. However, the utility of the pixelized images is quite low, due to the high perturbation noise required by differential privacy.

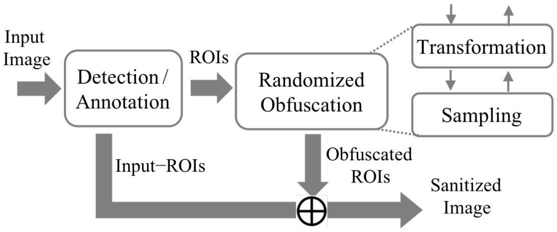

Figure 1. A Solution for Privacy-Preserving Image Sharing

This abstract presents a practical image obfuscation method to provide *provable* privacy guarantees, without compromising utility. Our method was first proposed in [4] to achieve metric privacy [1], a generalized notion based on differential privacy. The method guarantees similarity-based indistinguishability among images, providing privacy guarantees in worst-case scenarios and boosting the utility of the obfuscated image.

2. Method

Overview. Figure 1 depicts the solution of sharing obfuscated image data. Sensitive ROIs in an image are extracted via face/text detection or data owner annotations, and obfuscated by a randomized algorithm. The obfuscation method involves two steps: transformation and sampling. The *transformation* steps maps an input ROI to a feature vector, and privacy is achieved by *sampling* in a metric space; the privacy-enhanced vector will the go through the inverse transform, resulting in the obfuscated ROI.

Metric Privacy. While the standard *differential privacy* [2] is a rigorous privacy notion, it is only applicable to sanitizing aggregate statistics. The authors of [1] extended the principle of differential privacy and proposed *metric privacy* to protect secrets in an arbitrary domain $\mathcal{X}$. Formally, a private mechanism K (in Equation 1) relies on a *distance* metric $d_{\mathcal{X}}$ between secrets and guarantees a level of indistinguishability proportional of the distance. For an adversary who observes the output space, e.g., Z, it is challenging to infer the exact input, thus the privacy of the input is protected.

$$K(x)(Z) \leq e^{\epsilon \times d_{\mathcal{X}}(x,x')} K(x')(Z), \quad \forall \text{ output } Z \qquad (1)$$

The guarantee of metric privacy relies on ϵ, as defined in Equation 1: lower ϵ indicates higher indistinguishability, hence stronger privacy. When Hamming distance is adopted, it has been shown [1] that metric privacy is equivalent to ϵ-differential privacy.

Transformation and Sampling. To achieve metric privacy, our obfuscation method transforms an ROI to a feature vector, and randomly samples a privacy-enhanced vector following specialized probability distributions. Without speculating on feature extraction, we adopted Singular Value Decomposition (SVD), which has been applied to preserve image perceptual similarity [5]. Only k most significant singular values are preserved and the rest are set to zeros. Since SVD can be carried out by the adversary, the sampling distributions need to provide plausible deniability for the input singular values. As a result, indistinguishability can be guaranteed even among similar images, e.g., those with the same singular matrices but slightly different singular values. We described in [4] one specific sampling distribution and formally prove its metric privacy guarantees, with Euclidean distance in the transformed space. Challenges for sampling implementation as well as proposed solutions can be found in the conference paper [4].

3. Results and Discussions

To illustrate the trade-off between privacy and utility, we obfuscated images in the *AT&T* faces database with various privacy levels, i.e., ϵ values, in Figure 2. As the privacy is relaxed in Figure 2, e.g., Row 2 vs. 3 and 4, obfuscated images exhibit more resemblance to the originals, indicating higher utility. We compare our method with the pixelization method with standard differential privacy in [3], i.e., Row 4 vs. 5. With the same ϵ value, our method results in higher utility than [3]. A complete set of empirical results can be found in the conference paper [4].

One limitation of the proposed obfuscation is the sharp reduction in qualitative utility when providing stronger privacy, e.g., from $\epsilon = 0.3$ to 0.1 in Figure 2. We will consider other feature extraction methods for *transformation* and adopt post-processing methods, in hopes of achieving a smoother trade-off between privacy and utility. Another future work direction is to extend our approach to obfuscate color images, while addressing new sampling challenges, as well as utility loss, due to increased dimensionality.

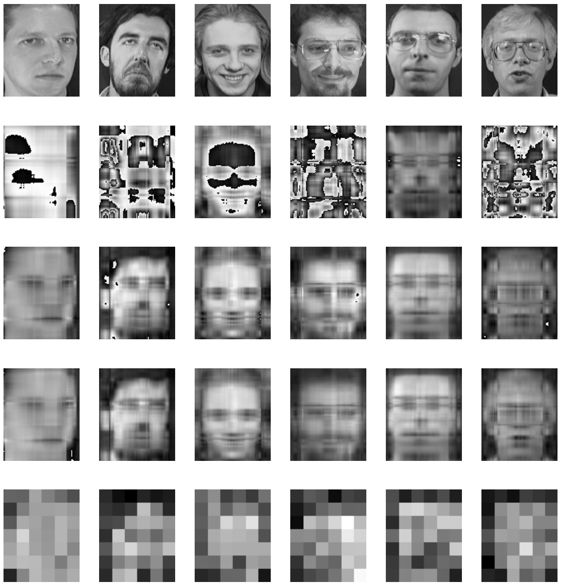

Figure 2. Example images and correponding obfuscation. Row 1 - original *AT&T* images; Row 2 - our obfuscation with $\epsilon = 0.1$; Row 3 - our obfuscation with $\epsilon = 0.3$; Row 4 - our obfuscation with $\epsilon = 1$; Row 5 - images obfuscated by [3] with $\epsilon = 1$.

References

[1] Konstantinos Chatzikokolakis, Miguel E. Andrés, Nicolás Emilio Bordenabe, and Catuscia Palamidessi. Broadening the scope of differential privacy using metrics. In Emiliano De Cristofaro and Matthew Wright, editors, *Privacy Enhancing Technologies*, pages 82–102, Berlin, Heidelberg, 2013. Springer Berlin Heidelberg. 1, 2

[2] Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. *Calibrating Noise to Sensitivity in Private Data Analysis*, pages 265–284. Springer Berlin Heidelberg, Berlin, Heidelberg, 2006. 1

[3] Liyue Fan. Image pixelization with differential privacy. In Florian Kerschbaum and Stefano Paraboschi, editors, *Data and Applications Security and Privacy XXXII*, pages 148–162, Cham, 2018. Springer International Publishing. 1, 2

[4] Liyue Fan. Practical image obfuscation with quantifiable privacy. In *Proceedings of the IEEE International Conference on Multimedia and Expo*, pages –. To Appear, 2019. 1, 2

[5] S. S. Kozat, R. Venkatesan, and M. K. Mihcak. Robust perceptual image hashing via matrix invariants. In *Image Processing, 2004. ICIP '04. 2004 International Conference on*, volume 5, pages 3443–3446 Vol. 5, Oct 2004. 2

[6] Richard McPherson, Reza Shokri, and Vitaly Shmatikov. Defeating image obfuscation with deep learning. *CoRR*, abs/1609.00408, 2016. 1

[7] E. M. Newton, L. Sweeney, and B. Malin. Preserving privacy by de-identifying face images. *IEEE Transactions on Knowledge and Data Engineering*, 17(2):232–243, Feb 2005. 1

[8] Zhongzheng Ren, Yong Jae Lee, and Michael S Ryoo. Learning to anonymize faces for privacy preserving action detection. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 620–636, 2018. 1

[9] Qianru Sun, Liqian Ma, Seong Joon Oh, Luc Van Gool, Bernt Schiele, and Mario Fritz. Natural and effective obfuscation by head inpainting. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018. 1