FaceWard: Face Anonymization in Group Photos

Sena Kilinç
University of Luxembourg
Luxembourg

Manasvi Ponaka
University of Luxembourg
Luxembourg

Luis A. Leiva
University of Luxembourg
Luxembourg

ABSTRACT

Sharing photos on social media and messaging services often result in a vast amount of personal data being made public online. As a result, it has become increasingly vital to devise measures that ensure privacy protection, especially for people who want to maintain social boundaries by hiding their faces in group photos. In this paper, we propose FaceWard, an automatic system for face anonymization of people different from a target person. FaceWard is based on a pattern matching algorithm and supports different anonymization policies such as blur or smiley overlays. Taken together, FaceWard yields a very efficient solution, eliminating the need for computationally expensive training of complex machine learning models, thus offering a practical trade-off between prediction accuracy and data availability. FaceWard is publicly available as open source software.

CCS CONCEPTS

• **Human-centered computing** → **Ubiquitous and mobile computing systems and tools**; • **Security and privacy** → *Privacy protections*; • **Information systems** → Multimedia content creation.

KEYWORDS

Anonymization; Obfuscation; Privacy

ACM Reference Format:
Sena Kilinç, Manasvi Ponaka, and Luis A. Leiva. 2023. FaceWard: Face Anonymization in Group Photos. In *25th International Conference on Mobile Human-Computer Interaction (MobileHCI '23 Companion), September 26–29, 2023, Athens, Greece.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3565066.3608249

1 INTRODUCTION

As face recognition technology continues to evolve, one unintended consequence is the emergence of significant privacy concerns. Namely, activities we do every day like sharing photos on social media often result in a vast amount of biometrics data being made public online. Given the ease of data accessibility in today's digital world, there is a growing risk of malicious exploitation of this sensitive data. Indeed, the main biometric personal identifier in images and videos is the human face [2]. As a result, it has become increasingly vital to devise measures that ensure the protection of such personal data.

MobileHCI '23 Companion, September 26–29, 2023, Athens, Greece

ACM ISBN 978-1-4503-9924-1/23/09.
https://doi.org/10.1145/3565066.3608249

Moreover, with stricter laws coming into play to regulate the use and collection of personal information, we find ourselves amidst heated debates about data privacy and daily life exposure. Particularly, the capture of user data through security cameras, autonomous vehicles, academic datasets, etc. is under scrutiny in many countries. A promising strategy to address this privacy issue involves image obfuscation. Traditionally, image obfuscation involves techniques such as blurring, pixelating, or blocking (masking) some parts such as the user eyes. Also, previous work has focused on obfuscating all faces in an image indiscriminately. Our approach, FaceWard, proposes a pivot from this convention. *We propose an automatic system for face anonymization of people different from a target person.* FaceWard takes as input a group picture and generates as output the same picture with all non-target faces obfuscated. FaceWard supports different anonymization policies, that can be either set as a default anonymization policy for all images that the user uploads or be selected on a per-image basis.

Instead of using sophisticated deep learning techniques, we propose a training-free and walk-up-and-use approach. Concretely, we employ a classic feature extraction algorithm followed by pattern matching to identify and anonymize faces not matching that of a target user. This approach yields a very efficient solution, eliminating the need for computationally expensive training of complex machine learning models, thus offering a practical trade-off between algorithmic accuracy and data availability. Ultimately, this research can serve as a launchpad for future explorations, potentially extending our application to videos and other face anonymizing challenges.

2 RELATED WORK

Face detection and related tracking technologies represent a vigorous field of study in many research areas, especially in computer vision and security. However, the concept of face anonymization, while not new, can be seen in a new light when considering the application put forth by this line of research.

Early work used methods based on the Haar-Cascade detector [8], which is an implementation of the popular Viola-Jones algorithm [10]. As we explain later, this is the foundation of FaceWard. More recent work has proposed deep learning approaches based on Convolutional Neural Networks (CNNs) [4]. For example, Kaipeng et al. [4] proposed an adaptive depth-wise CNN for automatically learning discriminating features better that outperformed many competing approaches. Their architecture paved way to many other approaches [6, 11]. More recently, DeepFace [9] focused on learning compact face representations from CNNs and achieved a detection error on par with human perception in recognizing faces. Other methods such as CLEANIR [1], DeepPrivacy [3], and DeepFakes [5] use deep generative models for de-identification and obfuscation. The downside of deep learning approaches is that they

require a large amount of training data and, also importantly, they are not easy to deploy in practice.

At its core, every face recognition technology relies on some kind of feature extraction. A sensible approach that has been shown to perform well in many scenarios is that of using Haar Cascades for automatic face feature extraction followed by pattern matching to identify the faces [12]. In a nutshell, once a face region is identified, facial features such as eyes, nose, and mouth are automatically extracted. This results in the creation of a face pattern, a type of template that encapsulates unique features of the user's face, that can be used to query a database of know faces. Then, if the query fails to find a satisfactory match, that face is deemed 'unknown' and thus can be anonymized, thereby ensuring the privacy of individuals not included in the reference database. Our approach leverages this well-established technique, pivoting thus from traditional machine learning methodologies and enabling an efficient and easy-to-use face anonymization system.

3 SYSTEM DESCRIPTION

Figure 1 illustrates the system workflow of FaceWard. When the app is accessed for the first time, it does not allow the user to upload any photo for anonymization, as the system has no information about the user's face yet. Instead, it is required to store some user photos first. This is done via video recording, in order to extract multiple frames that can account for different head orientations and poses. FaceWard automatically selects the three most dissimilar user faces for storage. Afterwards, the user can upload any photo to be anonymized and apply one of the following anonymization policies: blur, pixelate, blackbox, and smiley.

The backbone of our application lies in the above-mentioned automatic feature extraction algorithm followed by pattern matching. It is designed to (1) recognize the user's face in an uploaded image, (2) distinguish it from any other faces present in the image, and (3) apply an anoymization policy to them. And it does so efficiently and accurately. In the following we describe our implementation in detail.

3.1 Feature extraction

Our face anonymization algorithm operates by comparing selected patterns of the uploaded image with the stored reference images obtained from the three most dissimilar frames of the initial user video. These patterns consist of various features that characterize the user's face. For example, they might include specific arrangements of facial landmarks such as the eyes, nose, and mouth, as well as more subtle details such as the texture and skin color. By focusing on such details, FaceWard can handle a wide range of head orientations and facial expressions, making it robust against variations in pose and lighting conditions.

The feature extraction algorithm then uses cross-correlation, a principle most often found in signal processing which consists in a sliding window that evaluates pixel-level connectivity across the uploaded image. For each position of this window, a similarity score is calculated. The positions with the highest scores indicate potential matches for a face. After this, a list of the most probable bounding boxes of each face in an image is obtained.

This process has a couple of significant advantages for real-time image processing. First and foremost, it offers considerable flexibility. It can detect faces even if they appear in different scales or orientations within the uploaded image. In addition, it is quite efficient in terms of computational resources. It can process images quickly, delivering results almost instantly. This quickness is a crucial feature, setting our algorithm apart from more sophisticated methods such as deep learning models. While these models can be powerful, they often require a great deal of computational resources, including training time and a large amount of representative data. Our approach prioritizes efficiency and speed without compromising the quality of face detection.

3.2 Pattern matching

We implemented the Local Binary Patterns (LBP) algorithm [7] to recognize the target user's face in the image. First, each detected bounding box is divided into smaller cells and, for each pixel in a cell, an 8-connected components comparison is performed. This gives an 8-digit binary number. Then, a histogram of the frequency of each number is computed over the cell. Next, the histograms of all cells are normalized and concatenated. This process gives a feature vector for each face's bounding box. Finally, to find the closest face, the feature vectors are compared using k-nearest neighbors (we use $k = 3$ sample faces from the target user) with the χ^2 distance as dissimilarity metric. The smallest distance is chosen as the final classification, which represents the target person in the image. The remaining faces are the ones to be anonymized.

3.3 Anonymization policies

In Figure 2 we show examples of the different anonymization policies, depicted as "styles" in the app, to make the concept more palatable to end users. The anonymization styles are implemented over each of the non-target bounding boxes, as follows:

- **Blur:** Applies a low-pass filter on the bounding box region.
- **Pixelate:** Divides the bounding box region into blocks, gets the dominant color in each block, and replace the color of all pixels in the block by such a dominant color.
- **Blackbox:** Paints in black color all pixels in the bounding box region.
- **Smiley:** Overlays a smiley face icon and scales it proportionally to the size of each bounding box.

3.4 Screenshots

The following figures illustrate the main features of FaceWard. Figure 3 shows the welcome screen at the left and the home screen at the right. The welcome screen is only visible to first-time users, which guides them through the app and explain the next steps. The home screen is the entry point of FaceWard, shown every time the user opens the app.

Figure 4 shows the video recording screen. Instead of uploading individual photos of the user's face to capture different head orientation and poses, the user is asked to record a video so that the app can extract multiple frames automatically, in the background. Afterwards, the user can upload any photo for anonymization.

Finally, Figure 5 shows the photo upload screen. Here, provided that the app has already stored face samples of the user, it is possible

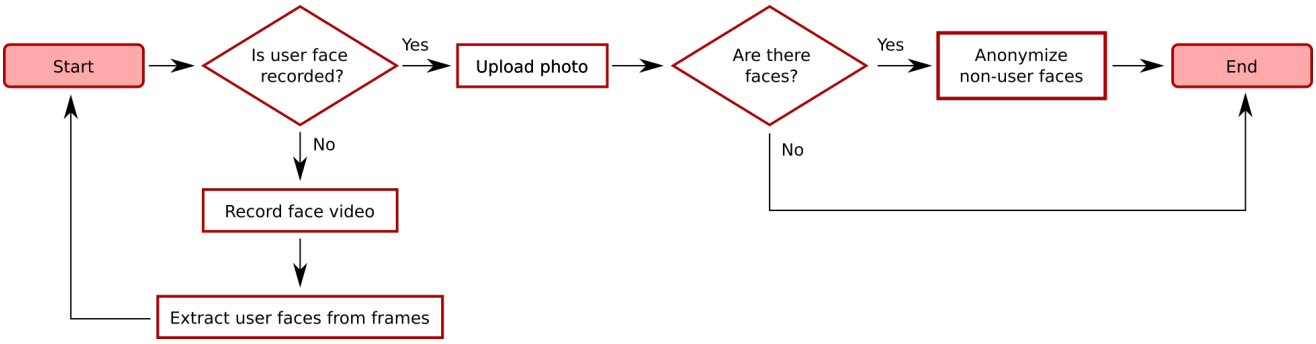

Figure 1: System workflow in FaceWard.

Figure 2: Examples of anonymization policies, from left to right: blur, pixelate, blackbox, smiley.

to submit any picture stored in the users' device. Otherwise, if the app has no knowledge yet about the user face, the user is asked to record a video first (see Figure 1). Finally, the user is asked to indicate the desired anonymization style to apply, after which the app will render the anonymized image in place. This way, the user can choose to either save the image or apply another anonymization style.

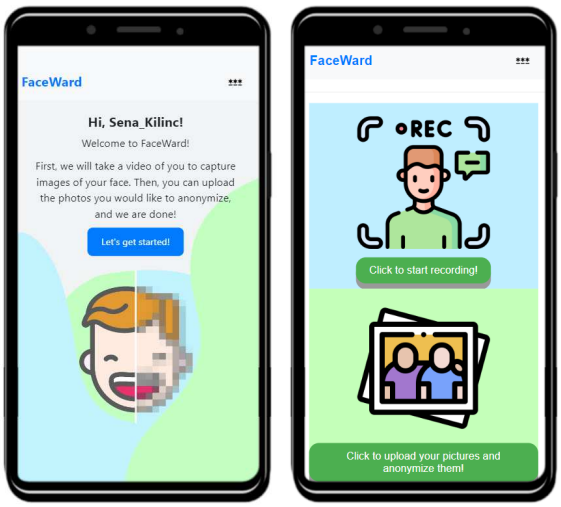

Figure 3: First-time usage screen (left) and welcome screen (right).

4 CONCLUSION AND FUTURE WORK

We have presented FaceWard, a simple yet efficient system for face anonymization of people different from a target person in group photos. By querying a small set of known user faces, FaceWard preserves other users' privacy by anonymizing unknown faces detected in the images taken by the user. Future work will consider multi-target identification (instead of our current single-target approach), so that it would be possible to preserve more than one target user, and a user evaluation comparing state-of-the-art approaches, such as the deep learning models mentioned in Section 2. FaceWard is publicly available as open-source software at https://github.com/senakilinc/FaceWard.

ACKNOWLEDGMENTS

This work was supported by the Horizon 2020 FET program of the European Union (grant CHIST-ERA-20-BCI-001) and the European Innovation Council Pathfinder program (SYMBIOTIK project, grant 101071147).

REFERENCES

[1] Durkhyun Cho, Jin Han Lee, and Il Hong Suh. 2020. CLEANIR: Controllable Attribute-Preserving Natural Identity Remover. *Applied Sciences* 10 (2020). Issue 3. https://doi.org/10.3390/app10031120

[2] J. Dietlmeier, J. Antony, K. McGuinness, and N. E. O'Connor. 2021. How important are faces for person re-identification?. In *Proc. International Conference on Pattern Recognition (ICPR)*. 6912–6919. https://doi.org/10.1109/ICPR48806.2021.9412340

[3] Håkon Hukkelås, Rudolf Mester, and Frank Lindseth. 2019. DeepPrivacy: A Generative Adversarial Network for Face Anonymization. In *Proc. International Symposium on Visual Computing (ISVC)*. 565–578. https://doi.org/10.1007/978-3-030-33720-9_44

[4] Zhang Kaipeng, Zhang Zhanpeng, Li Zhifeng, and Qiao Yu. 2016. Joint face detection and alignment using multi-task cascaded convolutional networks. *CoRR* abs/1604.02878 (2016).

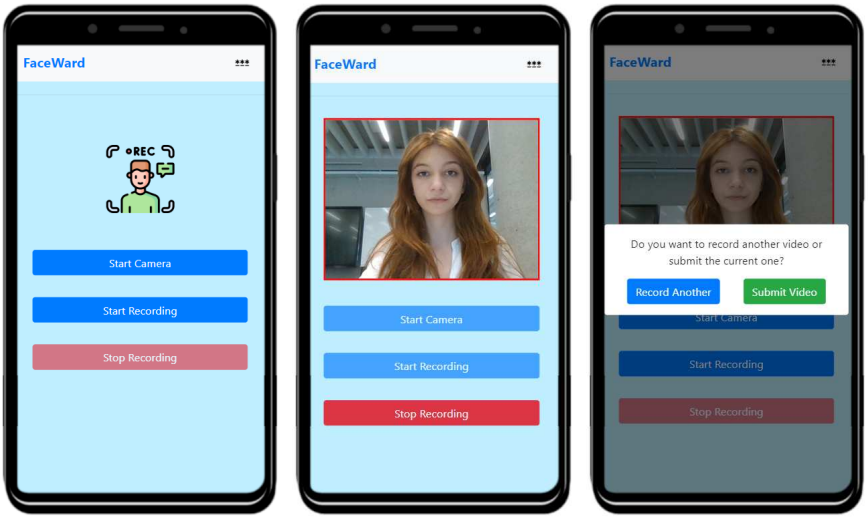

Figure 4: Video recording screen.

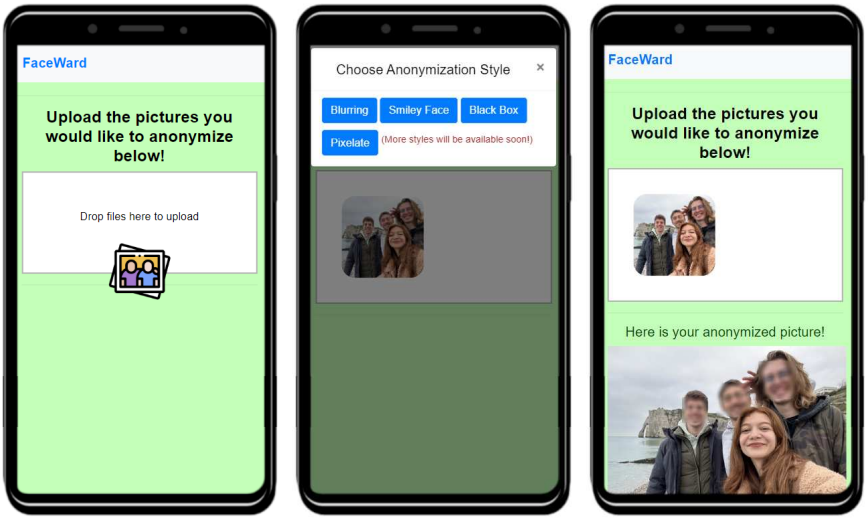

Figure 5: Photo upload screen (left), anonymization selection (middle), and final result (right).

[5] Mohamed Khamis, Habiba Farzand, Marija Mumm, and Karola Marky. 2022. DeepFakes for Privacy: Investigating the Effectiveness of State-of-the-Art Privacy-Enhancing Face Obfuscation Methods. In *Proc. International Conference on Advanced Visual Interfaces (AVI)*. https://doi.org/10.1145/3531073.3531125

[6] Mei Ma and Jianji Wang. 2018. Multi-View Face Detection and Landmark Localization Based on MTCNN. In *Proc. Chinese Automation Congress (CAC)*. 4200–4205. https://doi.org/10.1109/CAC.2018.8623535

[7] Timo Ojala, Matti Pietikäinen, and David Harwood. 1996. A comparative study of texture measures with classification based on featured distributions. *Pattern Recognition* 29, 1 (1996), 51–59. https://doi.org/10.1016/0031-3203(95)00067-4

[8] Rafael Padilla, Cicero Filho, and Marly Costa. 2012. Evaluation of Haar Cascade Classifiers for Face Detection. *World Academy of Science, Engineering and Technology* 64 (2012), 362–365.

[9] Yaniv Taigman, Ming Yang, Marc'Aurelio Ranzato, and Lior Wolf. 2014. DeepFace: Closing the Gap to Human-Level Performance in Face Verification. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE Computer Society, USA, 1701–1708. https://doi.org/10.1109/CVPR.2014.220

[10] Paul Viola and Michael Jones. 2001. Rapid object detection using a boosted cascade of simple features. In *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, Vol. 1. https://doi.org/10.1109/CVPR.2001.990517

[11] Meng Yuan, Seyed Yahya Nikouei, Alem Fitwi, Yu Chen, and Yunxi Dong. 2020. Minor Privacy Protection Through Real-time Video Processing at the Edge. *CoRR* abs/2005.01178 (2020).

[12] Stefanos Zafeiriou, Cha Zhang, and Zhengyou Zhang. 2015. A survey on face detection in the wild: Past, present and future. *Computer Vision and Image Understanding* 138 (2015), 1–24. https://doi.org/10.1016/j.cviu.2015.03.015