Faceless Person Recognition; Privacy Implications in Social Media

Seong Joon Oh
joon@mpi-inf.mpg.de
Rodrigo Benenson
benenson@mpi-inf.mpg.de
Mario Fritz
mfritz@mpi-inf.mpg.de
Bernt Schiele
schiele@mpi-inf.mpg.de

Computer Vision and Multimodal Computing
Max Planck Institute for Informatics, Germany

Person A training samples. Is this person A ?

Figure 1: An illustration of one of the scenarios considered: can a vision system recognise that the person in the right image is the same as the tagged person in the left images, even when the head is obfuscated?

Fully visible Gaussian blur Black fill-in White fill-in

Figure 2: Obfuscation types considered.

1 Introduction

With the growth of the internet, more and more people share and disseminate large amounts of personal data online. It is clear that visual data contains private information, yet the privacy implications of this data dissemination are unclear, even for computer vision experts. We are aiming for a transparent and quantifiable understanding of the loss in privacy incurred by sharing personal data online.

An important component to extract maximal information out of visual data in social networks is to fuse different data and provide a joint analysis. We propose our new Faceless Person Recogniser which not only reasons about individual images with cues from face, body, and contextual regions, but uses graph inference to deduce identities in a group of non-tagged images. Our contributions are:

- Discuss dimensions that affect the privacy of online photos, and define a set of scenarios to study the question of privacy loss when such images are aggregated and processed by a vision system.
- Propose our new Faceless Person Recogniser, which uses convnet features in a graphical model for joint inference over identities.
- Study the interplay and effectiveness of obfuscation techniques with regard of our vision system.

2 Privacy Scenarios

We consider four different dimensions that affect how hard or easy it is to recognise a user:

- Number of tagged heads. (Table 1)
- Obfuscation type. (Figure 2)
- Amount of obfuscation. (Table 1)
- Domain shift. (Tagged and non-tagged images can be either "across events" or "within events". Figure 3)

Based on this, we propose to consider privacy scenarios, S_0, S_1^τ, S_2 and S_3 in table 1, with scenario applied to "across events" and "within events".

Table 1: Privacy scenarios considered. Each row in the table can be applied for the "across events" and "within events" case, and over different obfuscation types. In scenario S_1^τ, $\tau \in \{1.25, 2.5, 5, 10\}$.

	Brief description	#Tagged	Obfuscation
S_0	Privacy indifferent	10	0%
S_1^τ	Some of my images tagged	τ	0%
S_2	One non-tagged head obfuscated	10	1 instance
S_3	All my heads obfuscated	10	100%

3 Experimental Setup

We investigate the scenarios proposed above through a set of controlled experiments on a recently introduced social media dataset (PIPA).

PIPA dataset. The PIPA dataset [4] consists of annotated social media photos on Flickr. It contains ∼40k images over ∼2k identities, and captures subjects appearing in diverse social groups (e.g. friends, colleagues, family) and events (e.g. conference, vacation, wedding).

Domain shift. The "Original" split proposed by [4] has tagged and non-tagged instances of people in the same events, while the "Day" split proposed by [2] splits the tagged and non-tagged instances according to events, clothing, and scene changes. We use the "Original" split as a proxy for the "within events" case, and the "Day" split for "across events".

Figure 3: PIPA dataset examples of person X. Vertically, upper half shows tagged images of X in "within events", and lower half corresponds to non-tagged images of X to be recognised. Analogous for horizontal splits in "across events".

4 Faceless Recognition System

We introduce the Faceless Recognition System to study the effectiveness of privacy protective measures in §2. Our system does joint recognition employing a conditional random field (CRF) model:

$$\arg\max_Y \frac{1}{|V|}\sum_{i\in V}\phi_\theta(Y_i|X_i) + \frac{\alpha}{|E|}\sum_{(i,j)\in E} 1_{[Y_i=Y_j]}\psi_{\tilde{\theta}}(X_i, X_j) \qquad (1)$$

with observations X_i, identities Y_i. $1_{[\cdot]}$ is the indicator function, and $\alpha > 0$ controls the unary-pairwise balance. For full details and analysis of methods, refer to our ICCV'15 [2] and ECCV'16 [3] papers.

4.1 Unary ϕ_θ: ICCV'15 [2]

We build our unary on each node $i \in V$ upon our previous state of the art person recogniser, `naeil` [2]. `naeil` extracts 17 identity relevant cues from 5 different regions around the person (figure 4). In particular, the method does not require a visible face for recognition; body and scene context can give useful information for identifying a person. `naeil` is shown to be robust to decreasing number of tagged examples.

A similar approach, combining cues from multiple context regions, is proposed in [4]. Unlike [4], we show that simple AlexNet cues from fixed context regions can already achieve a better performance than the one with specialised face features and pose estimation technique.

Figure 4: Single person recognition proposed in [2]

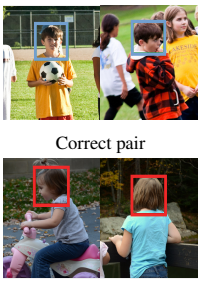

Figure 5: Matching in social media.

4.2 Pairwise $\psi_{\tilde{\theta}}$.

By adding pairwise terms over the unaries, we expect the system to propagate predictions across nodes. When a unary prediction is weak (e.g. obfuscated head), the system aggregates information from connected nodes with possibly stronger predictions (e.g. visible face), and thus deduce the query identity. Our pairwise term $\psi_{\tilde{\theta}}$ is a siamese network. Note that matching problem in social media setting is challenging (figure 5).

5 Test set results & analysis

We evaluate our Faceless Recognition System on the PIPA test set. The main results are summarised in figures 6 and 7. We organize the results along the same privacy sensitive dimensions that we defined in §2. For extended analysis and discussion, we refer to [3].

Number of tagged heads. Figure 6 shows that even with only 1.25 tagged photos per person on average, the system can recognise 73× chance level for "wihin events". We see that even few tags provide a threat for privacy and thus users concerned with their privacy should avoid having (any of) their photos tagged.

Obfuscation type. Figure 7 shows that, from higher protection to lower protection, we have Black ≈ White > Blur > Visible. Albeit blurring does provide some protection, recognition rate still remains high: ∼100× and ∼20× chance level for "within/across events", respectively.

Amount of obfuscation. We cover three scenarios: every head fully visible (S_1), only the test head obfuscated (S_2), and every head fully obfuscated (S_3). Figure 7 shows that within events obfuscating either one (S_2) or all (S_3) heads is not very effective, compared to the across events case, where one can see larger drops for $S_1 \rightarrow S_2$ and $S_2 \rightarrow S_3$. We conclude that within events head obfuscation has only limited effectiveness, across events only blacking out all heads seems truly effective (S_3 black).

Domain shift. In all scenarios, the recognition accuracy is significantly worse in the across events case than within events. For a user, it is a better privacy policy to make sure no tagged heads exist for the same event, than blacking out all his heads in the event.

6 Discussion & Conclusion

Within the limitation of any study based on public data, we believe the results presented here are a fresh view on the capabilities of machine learning to recognise people in social media under adversarial condition. From a privacy perspective, the results presented here should raise concern. It is very probable that undisclosed systems similar to the ones described here already operate online. We believe it is the responsibility of the computer vision community to quantify, and disseminate the privacy implications of the images users share online. This work is a first step in this direction. We conclude by discussing some future challenges and directions on privacy implications of social visual media.

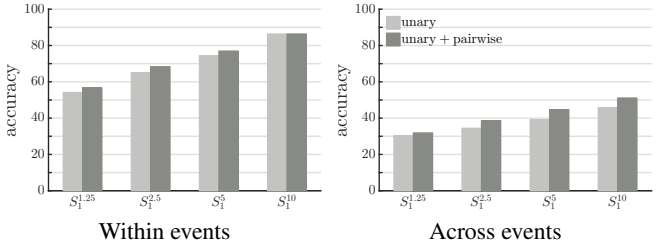

Figure 6: Impact of number of tagged examples: $S_1^{1.25}$, $S_1^{2.5}$, S_1^{5}, and S_1^{10}.

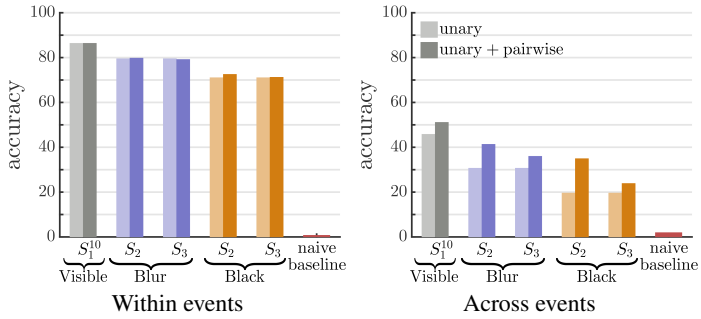

Figure 7: Co-recognition results for scenarios S_1^{10}, S_2, and S_3 with black fill-in and Gaussian blur obfuscations (white fill-in match black results).

Lower bound on privacy threat. The current results focused singularly on the photo content itself and therefore a lower bound of the privacy implication of posting such photos. It remains as future work to explore an integrated system that will also exploit the images' meta-data (timestamp, geolocation, camera identifier, related user comments, etc.). In the context of the era of "selfie" photos, meta-data can be as effective as head tags. Younger users also tend to cross-post across multiple social media, and make a larger use of video (e.g. Vine). Using these data-form will require developing new techniques.

Training and test data bounds. The performance of recent techniques of feature learning and inference are strongly coupled with the amount of available training data. Many state of the art person recognition systems rely on undisclosed training data in the order of millions of training samples. Similarly, the evaluation of privacy issues in social networks requires access to sensitive data, which is often not available to the public research community (for good reasons [1]). The used PIPA dataset [4] serves as good proxy, but has its limitations. It is an emerging challenge to keep representative data in the public domain in order to model privacy implications of social media and keep up with the rapidly evolving technology that is enabled by such sources.

From analysing to enabling. In this work, we focus on the analysis aspect of person recognition in social media. In the future, one would like to translate such analyses to actionable systems that enable users to control their privacy while still enabling communication via visual media exchanges.

Acknowledgements This research was supported by the German Research Foundation (DFG CRC 1223).

References

[1] Arvind Narayanan and Vitaly Shmatikov. Myths and fallacies of personally identifiable information. *Communications of the ACM*, 2010.

[2] Seong Joon Oh, Rodrigo Benenson, Mario Fritz, and Bernt Schiele. Person recognition in personal photo collections. In *ICCV*, 2015.

[3] Seong Joon Oh, Rodrigo Benenson, Mario Fritz, and Bernt Schiele. Faceless person recognition; privacy implications in social media. In *ECCV*, 2016. to appear.

[4] Ning Zhang, Manohar Paluri, Yaniv Taigman, Rob Fergus, and Lubomir Bourdev. Beyond frontal faces: Improving person recognition using multiple cues. In *CVPR*, 2015.