Anonymizing Egocentric Videos

Daksh Thapar, Aditya Nigam
Indian Institute of Technology Mandi
dakshthapar.github.io, faculty.iitmandi.ac.in/~aditya

Chetan Arora
Indian Institute of Technology Delhi
www.cse.iitd.ac.in/~chetan

1. Introduction

In the paper, we briefly discussed the datasets that have been used to validate our proposed transformation framework. Then we have shown the recognition capability of EgoIdNet under both closed-set (wearers are known and trained for during training) and open-set (wearers are unseen during training) scenarios as shown in Table 1.

Table 1: Comparative analysis of our system with [6] and [7] for wearer recognition in egocentric videos. CA, and EER denote the classification accuracy, and Equal Error Rate respectively in percentage. Higher CA and lower EER is better.

Dataset	Closed Set Analysis				Open Set Analysis			
	EgoIDNet		[6]		EgoIDNet		[6]	
	CA	EER	CA	EER	EER	CRR	EER	CRR
FPSI	82.4	18.76	82.0	19.71	–	–	–	–
DB-01	99.1	2.47	99.2	2.79	5.87	86.33	6.43	83.67
DB-02	97.6	3.54	97.3	3.81	7.67	84.28	8.23	82.77
DB-03	99.0	4.12	98.7	4.35	6.52	83.46	9.39	80.56
	EgoIDNet		[7]		EgoIDNet		[7]	
EPIC	71.21	12.46	71.04	12.32	14.32	59.67	15.28	55.06

Finally we have shown the comparative performance of our proposed transformation framework with noise model as shown in Figure 1

In this supplementary material, We first provide more details about the datasets used. We then provide the performance values of the comparative analysis and also show the the anonymizing performance of our proposed approach on our own EgoIDNet framework. Finally we have attached a video demo for qualitative analysis of the proposed approach.

2. Dataset Specifications

Table 2: Dataset Specifications. Every clip of 4 seconds duration is one data input.

Dataset	Subjects	Training Data	Testing Data	Genuine Matching	Imposter Matching
FPSI	6	5693	7268	5255000	26275000
EVPR	32	1652	1708	7680	53760
DB-01	31	2012	2384	30526	5625489
DB-02	31	2276	2345	35621	7526489
DB-03	31	4288	4729	50498	10256412

We validate the performance of our anonymizing strategy on the same two benchmark egocentric datasets as used by the attack techniques [5, 6, 7]:

1. **EPIC kitchens dataset [1] :** It consists of 55 hours of egocentric videos from 32 subjects, and contains 125 labeled activities performed by the subjects. As Thapar et al. [7] has validated the person recognition system on the five activities,

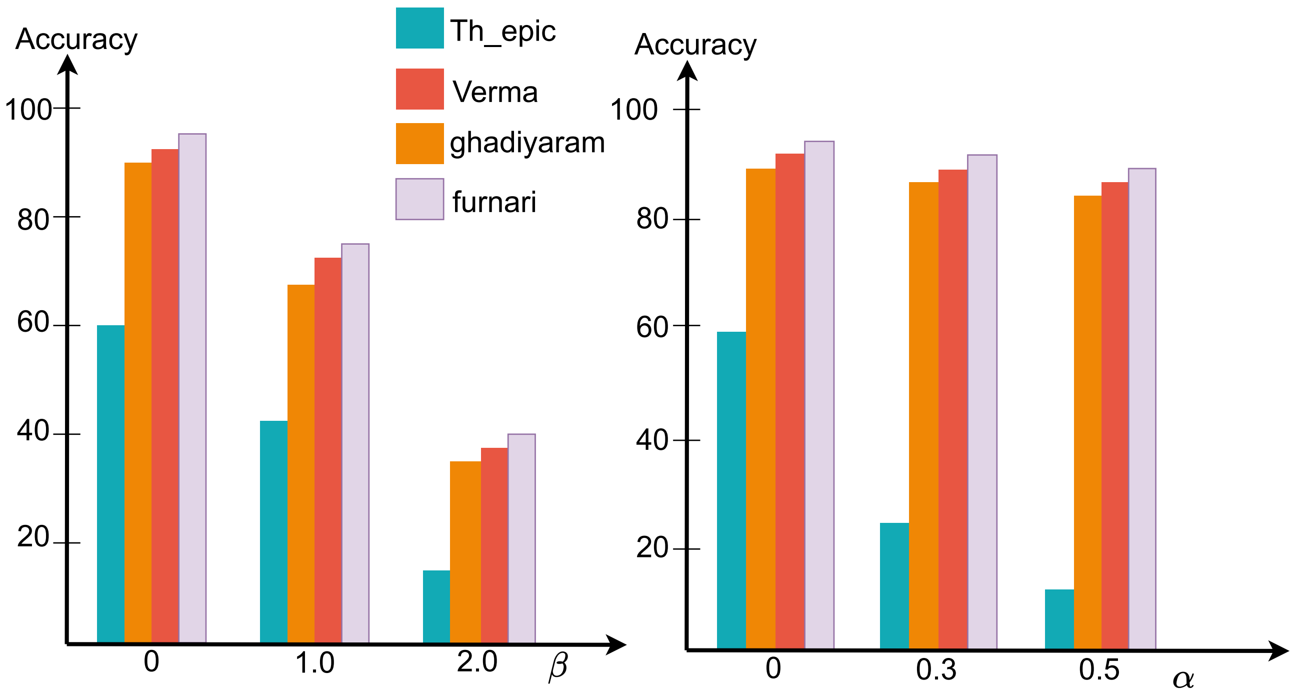

Figure 1: Comparison of anonymization performance and degradation in activity recognition on videos of "cut" activity in EPIC-Kitchens dataset after fine-tuning each model with the respective noise. The left plot corresponds to adding random noise with various level of β. The right plot is performance of proposed perturbation for various levels of α. See text for details of α and β. "Th_epic" show the wearer recognition by [7] (lower is better), whereas the other three bars show activity recognition performance of SOTA on the perturbed videos (higher values indicate no degradation and is better). Noise level 0 indicates no noise is added (i.e. original dataset).

Figure 2: The first two rows shows example frames from EPIC-Kitchens dataset. The last two rows shows example frames from FPSI dataset.

we have also chosen the same five activities viz *cut, mix, put, take*, and *wash*. Figure 2 shows the sample frames of EPIC Kitchens.

2. **The FPSI dataset [2]:** FPSI is a publicly available dataset consisting of video captured by 6 people wearing cameras mounted on their hat, and spending their day at Disney World Resort in Orlando, Florida. Figure 2 shows the sample frames of FPSI.

The detailed dataset specifications are shown in Table 2.

To validate the performance of our proposed EgoIDNet, we use another dataset IITMD-WFP [6] consisting of 3.1 hours of videos captured by 31 different subjects. The dataset has two different scenerios: indoor and outdoor, and refer to the respective datasets as DB-01 (indoor), and DB-02 (outdoor). The combined dataset is referred as DB-03.

3. Performance Comparison with Naive Noise Model

Table 3: The effect of adding noise with level β. We demonstrate the accuracy of each model for "cut" activity in EPIC Kitchens dataset. The 2nd column shows that noise level is 0, i.e. no noise is added (original dataset). "FT" indicates that the model has been fine-tuned over noisy samples.

Wearer Recognition

Model	$\beta = 0$	$\beta = 1.0$	$\beta = 2.0$
Th_epic[7]	59.9	42.7	12.5
Th_fpsi[7] (FT)	59.9	48.9	14.6
EgoIDNet (ours)	61.4	44.6	13.8
EgoIDNet (ours) (FT)	61.4	47.7	18.3

Activity Recognition

Model	$\beta = 0$	$\beta = 1.0$	$\beta = 2.0$
Verma[8]	92.4	68.7	32.6
Verma[8] (FT)	92.4	71.4	37.4
ghadiyaram[4]	91.5	67.2	30.6
ghadiyaram[4] (FT)	91.5	69.7	35.4
furnari[3]	92.9	70.6	33.3
furnari[3] (FT)	92.9	72.2	40.0

Table 4: Effect of adding proposed transformation at various blending levels α. We demonstrate the accuracy of each model on "cut" activity in EPIC Kitchens dataset. The 2nd column indicates blending level 0, i.e. no transformation (original dataset). "FT" indicates that the model has been finetuned over transformed dataset. (Transformation using blending level $\alpha = 0.3$ has been observed empirically to be the best)

Wearer Recognition

Model	$\alpha = 0$	$\alpha = 0.3$	$\alpha = 0.5$
Th_epic[7]	59.9	15.4	11.3
Th_fpsi[7] (FT)	59.9	17.8	14.5
EgoIDNet (ours)	61.4	11.3	7.6
EgoIDNet (ours) (FT)	61.4	12.4	9.3

Activity Recognition

Model	$\alpha = 0$	$\alpha = 0.3$	$\alpha = 0.5$
Verma[8]	92.4	85.2	72.3
Verma[8] (FT)	92.4	91.8	84.6
ghadiyaram[4]	91.5	83.3	71.8
ghadiyaram[4] (FT)	91.5	90.2	82.9
furnari[3]	92.9	87.5	73.1
furnari[3] (FT)	92.9	92.0	86.4

Table 3 shows the result of addition of varying amount of random noise, to the videos of "cut" activity from EPIC-Kitchens dataset. We compare the anonymization performance to block open set wearer recognition, where the bar corresponding to "Th_epic" show the wearer recognition by [7]. We see that with increasing level of random noise, the activity recognition performance drops at a similar rate as the wearer recognition, indicating significant interference of random noise in other video analysis task as well. Table 4 shows the result of adding proposed transformation at various blending scales. Here we can observe, with increasing amount of the proposed perturbation, wearer recognition falls while activity recognition falls marginally.

4. Qualitative Analysis

The attached video demo shows the qualitative analysis of the proposed system with respect to noise model. The perturbation levels shown in the video corresponds to the best anonymization performance corresponding to both the noise model and proposed framework. the pertubations leads to very bad visual output for the case of random noise perturbation. However, the visual output from our proposed framework is similar to the original egocentric video. The agenda for the supplementary video is to show the degradation of the noise model qualitatively.

References

[1] Dima Damen, Hazel Doughty, Giovanni Maria Farinella, Sanja Fidler, Antonino Furnari, Evangelos Kazakos, Davide Moltisanti, Jonathan Munro, Toby Perrett, Will Price, and Michael Wray. Scaling egocentric vision: The epic-kitchens dataset. In *European Conference on Computer Vision (ECCV)*, 2018. 1

[2] Alircza Fathi, Jessica K Hodgins, and James M Rehg. Social interactions: A first-person perspective. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1226–1233. IEEE, 2012. 2

[3] Antonino Furnari and Giovanni Maria Farinella. What would you expect? anticipating egocentric actions with rolling-unrolling lstms and modality attention. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 6252–6261, 2019. 3

[4] Deepti Ghadiyaram, Du Tran, and Dhruv Mahajan. Large-scale weakly-supervised pre-training for video action recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12046–12055, 2019. 3

[5] Yedid Hoshen and Shmuel Peleg. An egocentric look at video photographer identity. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4284–4292, 2016. 1

[6] Daksh Thapar, Aditya Nigam, and Chetan Arora. Is sharing of egocentric video giving away your biometric signature? In *The European Conference on Computer Vision (ECCV)*, August 2020. 1, 2

[7] Daksh Thapar, Aditya Nigam, and Chetan Arora. Recognizing camera wearer from hand gestures in egocentric videos. In *Proceedings of the 28th ACM International Conference on Multimedia*, MM ’20, page 2095–2103, New York, NY, USA, 2020. Association for Computing Machinery. 1, 2, 3

[8] Sagar Verma, Pravin Nagar, Divam Gupta, and Chetan Arora. Making third person techniques recognize first-person actions in egocentric videos. In *2018 25th IEEE International Conference on Image Processing (ICIP)*, pages 2301–2305. IEEE, 2018. 3