Designing a Camera for Privacy Preserving

Hajime Nagahara[1]
1 D3 Center, Osaka University
2-8, Yamadaoka, Suita, Osaka 565-0871, Japan
E-mail: nagahara@ids.osaka-u.ac.jp

Abstract *The widespread use of computer vision systems in our personal spaces has led to an increased consciousness of these systems' privacy and security risks. On the one hand, we want these systems to assist in our daily lives by understanding our surroundings, but on the other hand, we want them to do so without capturing any sensitive information. Towards this direction, we propose a method for designing a privacy-preserving camera that degrades the captured image quality by optics and sensor designs. However, the image still has information for downstream tasks. The proposed method models an imaging and image recognition pipeline as differential and neural models, then jointly optimizes the models in an end-to-end manner to find a good balance between image degradation and task accuracy. In this talk, I will talk about the concept of the method and show two examples of the downstream tasks: privacy-preserving identification for the human face and privacy-preserving action recognition. We confirmed that we realized captured images are visually hard to recognize identity by humans, but they maintain enough accuracies for identification and recognition by machine learning models.*

Keywords: Deep optics, Deep sensing, Privacy-preserving, Human identification, Action recognition

1. Introduction

People have been fascinated with creating computer vision (CV) systems that can see and interpret the world around them for many decades. In today's world, as this dream becomes a reality and such systems are developed in our personal spaces, there is an increased consciousness about "what" these systems see and "how" they interpret it. Nowadays, we want CV systems that protect our visual privacy without compromising the user experience. Therefore, there is growing interest in developing such CV systems that can prevent the camera system from obtaining detailed visual data that may contain sensitive information but allow it to capture valuable information to perform the CV task [1,2,3,4,5]. For a satisfactory user experience and strong privacy protection, a CV system must satisfy the following properties:

– Good target task accuracy. This is necessary for maintaining a good user experience. For example, a privacy-preserving face detection model must detect faces with high precision without revealing facial identity [2], a privacy-preserving pose estimation model must detect body key points without revealing the person's identity [4], and an action recognition model must recognize human actions without revealing their identity information [1,3].

– Strong privacy protection. Any privacy-preserving model, irrespective of the target task, must preserve common visual privacy attributes such as identity, gender, race, color, gait, etc.

In this talk, we propose a method for designing a privacy-preserving camera that degrades the captured image quality by optics and sensor designs. However, the image still has information for downstream tasks. The proposed method models an imaging and image recognition pipeline as differential and neural models, then jointly optimizes the models in an end-to-end manner to find a good balance between image degradation and task accuracy.

In this talk, I will talk about the concept of the method and show two examples of the downstream tasks: privacy-preserving identification of human faces [6,7] and privacy-preserving action recognition [8]. We confirmed that we realized captured images are visually hard to recognize identity by humans, but they maintain enough accuracies for identification and recognition by machine learning models.

References

[1] Wu, Z., Wang, Z., Wang, Z., Jin, H., “Towards privacy-preserving visual recognition via adversarial training: A pilot study”, European Conference on Computer Vision, pp. 606–624, 2018.

[2] Ryoo, M.S., Rothrock, B., Fleming, C., Yang, H.J., “ Privacy-preserving human activity recognition from extreme low resolution”, AAAI Conference on Artificial Intelligence, 2017.

[3] Srivastav, V., Gangi, A., Padoy, N., “Human pose estimation on privacy-preserving low-resolution depth images”, International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 583–591, Springer, 2019.

[4] Hinojosa, C., Niebles, J.C., Arguello, H., “Learning privacy-preserving optics for human pose estimation, IEEE/CVF International Conference on Computer Vision. pp. 2573–2582, 2021.

[5] Raval, N., Machanavajjhala, A., Cox, L.P., “Protecting visual secrets using adversarial nets”, IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 1329–1332, 2017.

[6] Nguyen Canh, T. and Nagahara, H., “Deep Compressive Sensing for Visual Privacy Protection in FlatCam Imaging”, ICCV workshop learning for computational imaging, Nov., Seoul Korea, 2019.

[7] Nguyen Canh, T, Thanh Ngo, T., Nagahara, H., “Human-Imperceptible Identification with Learnable Lensless Imaging”, IEEE access, Vol. 11, pp. 95724-95733, 2023.

[8] Kumawat, S. and Nagahara , H., “Privacy-Preserving Action Recognition via Motion Difference Quantization”, European Conference on Computer Vision, Oct., Tel Aviv, 2022.

Designing a Camera for Privacy Preserving

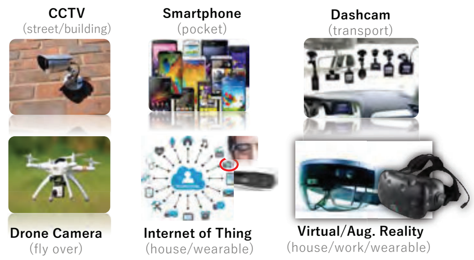
Hajime Nagahara
Osaka University

Visual Privacy

- **Privacy** is related to the *freedom from interference*, *state of being alone*, the right to keep *personal matters* and the *relationship secret* [1]
- **Visual Privacy** is the right to collect and use *personal visual information*[2]
  - Info (face, race, gender, clothes, license plate, etc.) that infers the personal identity

[1] https://www.merriam-webster.com/dictionary/privacy
[2] J. Shu, "A survey on Visual Privacy in Ubiquitous Computing," PhD qualifying examination survey., 2017.

Cameras Everywhere



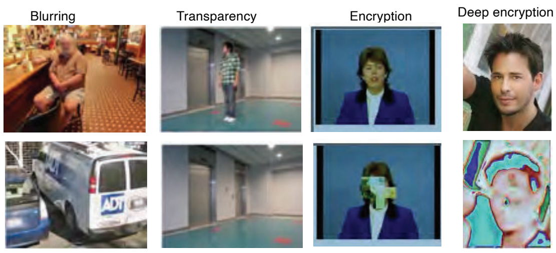
We are increasingly exposed!!!

Visual Privacy Violation

Uninformed Photography Surveillance Abuse | **Hacking Visual Data** | **Analysis Visual Data**

- Take picture of person/object wo. permission
- Access unauthorized sensitive visual contents
- Visual profiling, targeted marketing
- Criminal abuse
- Institutional abuse
- Discrimination & voyeurism

Existing privacy preserving by software



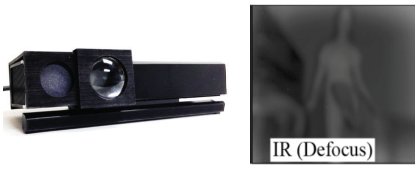
Google Street View | Babaguchi et al. | F. Peng *et al.* | Li et al. IoTDI21

Privacy sensing by thermal camera

Thermal imaging

F. Pittaluga, A. Zivkovic and S. J. Koppal, "Sensor-level privacy for thermal cameras," in IEEE International Conference on Computational Photography , 2016

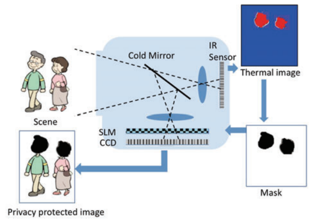
Privacy sensing by Defocus optics



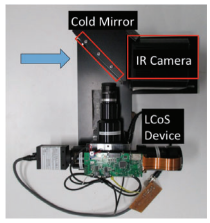
Defocused optics

F. Pittaluga and S. J. Koppal, "Privacy Preserving Optics for Miniature Vision Sensors," in IEEE International Conference on Computer Vision and Pattern Recognition, 2015.

Anonymous Camera (face masking camera)



Detecting facial region to make the mask and apply the optical mask in front of the sensor



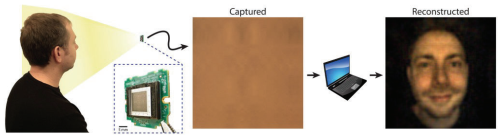
Prototype camera

[Zhang, ICPR2014]

Optically masked sensor image

Flatcam imaging

Face dataset
88 subjects, 24.112 images



- Training with simulated and real lensless data
- Initial reconstruction is required

Face detection with R-CNN and verification through classification CNN

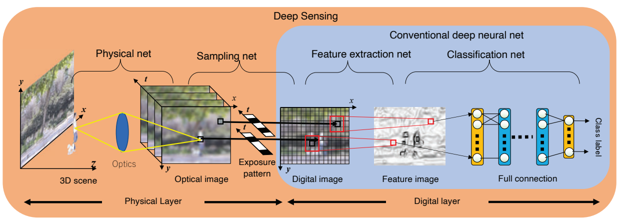
- *This system is not privacy as the initial reconstruction is required*
- *Not consider the privacy of classifier regards to mask*

J. Tan et al., Face detection and verification using lensless cameras, IEEE TCI 2019

Computational photography & deep sensing

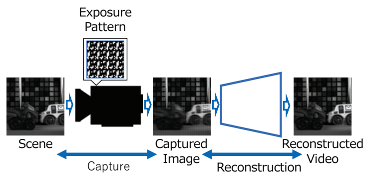
- Typical deep neural networks learn in the digital layers
- A physical layer may be facilitated by learning from data
- Optimizing the sensing hardware is possible by learning

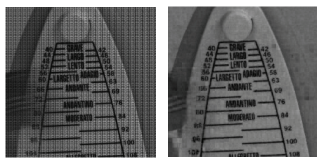
H30-R1挑戦的研究(萌芽)
R2-5 挑戦的研究(開拓)
R5-9 基盤研究S



Deep compressive video sensing

- Generating video (x16 frames) from a SINGLE shot
- The shutter patten and decoder are jointly optimized.



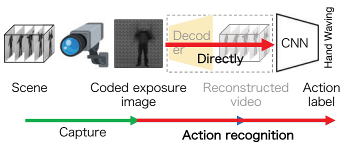


Captured images (3 frames)

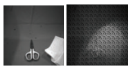
Reconstructed Video (3x16=48 frames)

Direct recognition from a coded image

- A coded exposure image has the temporal information.
- We can directly recognize the action from a single image.



Video (upper bound)	71.0
Coded exposure (Proposed)	72.0



Efficient image recognition
(1/16 Smaller data but similar accuracy)

Sensor Level Visual Privacy

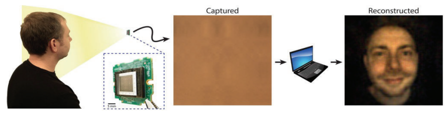
- It is considered to protect the hardware attack as well as digital attack via an internet.
- Existing privacy preserving camera is manually designed.
- The degradated captured image is also decrease downstream performance.
- We jointly train the optics and recognition model by using adversarial learning for balancing a privacy and utility.

Human-Imperceptible Identification With Learnable Lensless Imaging

Thuong Nguyen Canh, Trung Thanh Ngo, Hajime Nagahara
IEEE Access 2023

FlatCam imaging

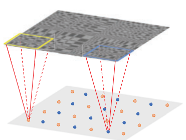
- It is also called lensless imaging.
- Photo Mask is placed in front of an imager.
- Reconstructed image is obtained from the blurred captured image.



J. Tan et al., Face detection and verification using lensless cameras, IEEE TCI 2019

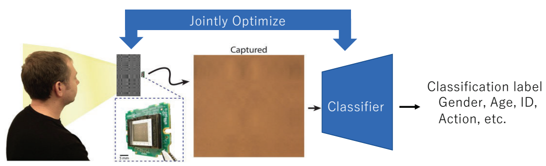
FlatCam imaging

- An intensity of each pixel is obtained as an integration of multiple rays though the mask modulation.
- The mask, e. g. M sequence, is uniquely modulated to each angle of the rays.
- The image is reconstructed by inverse processing of mask modulation.



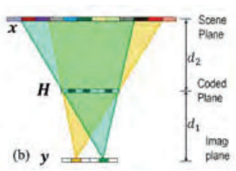
Proposed privacy preserving camera

- Using mask based FlatCam
- Jointly optimizing the mask pattern and utility classifier.
- Realizing good balance for degradeding the captured image for privacy preserving and maintaining classification accuracy.

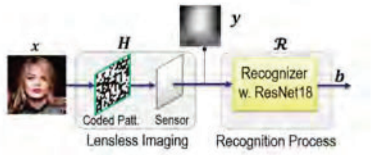


Modeling of proposed optics

- Modeling lensless imaging to convolution.
- We use ResNet for identification model.
- The coding mask pattern H and identification model R should be optimized simultaneously.

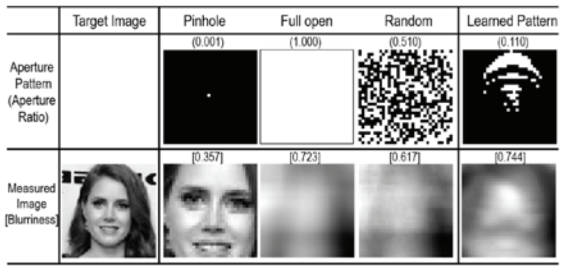


Lensless Imaging model



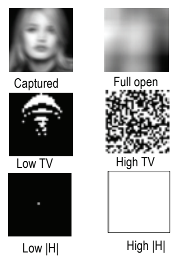
$$\mathcal{L} = \underbrace{\mathcal{L}_{\text{rec}}(\mathcal{R}(H * x), b)}_{\text{Utility loss}} + \alpha \times \underbrace{\mathcal{L}_{\text{hi}}}_{\text{Human imperception loss}}$$

Examples of coded pattern for FlatCam



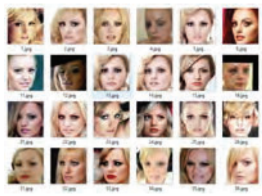
Human Imperceptible Loss

- Similarity loss: $\mathcal{L}_{\text{sim}} = \sum_i ||H * x_i - 1_m * x_i||_2^2.$
- Total variation loss: $\mathcal{L}_{\text{tv}} = -||\Delta_x H||_1 - ||\Delta_y H||_1$
- Invertibility loss: $\mathcal{L}_{\text{inv}} = -||H||_1.$
- RIP loss: $\mathcal{L}_{\text{rip}} = -\sum_i \frac{||H * x_i||_2^2}{||x_i||_2^2 + \epsilon}.$
  (easy to reconstruct or not)

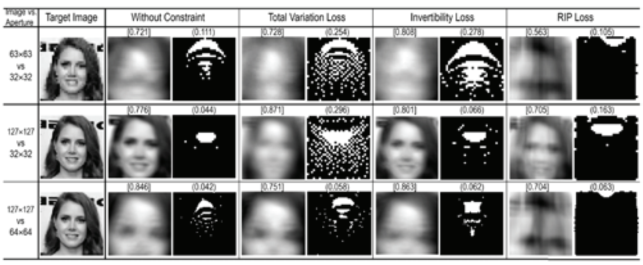


Face dataset

- Microsoft Celeb (MS-Celeb-1M): 10 million faces, 80k class
  - Align Dataset, 112x112.
- Train/Test subjects: ratio 95/5
  - 10 classes
- Others
  - Resize to 63x63
  - Mask of size 32x32
  - Test image is flipped, rotate to increase test size



Trained aperture pattens

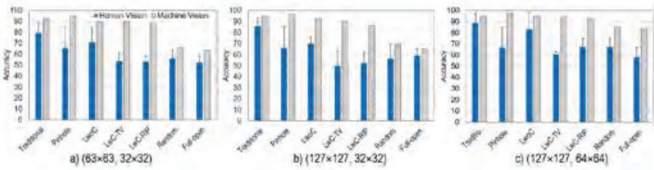


Recognition results

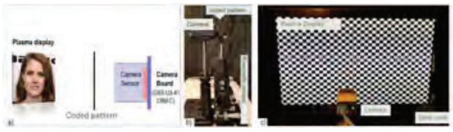
	Dataset	MS-Celeb			VGG-Face2			CASIA		
Patt.	Image size	63×63	127×127		63×63	127×127		63×63	127×127	
	Patt. size	32×32	32×32	64×64	32×32	32×32	64×64	32×32	32×32	64×64
	Coded ratio	1/4	1/16	1/4	1/4	1/16	1/4	1/4	1/16	1/4
Fix	Pinhole	**98.57**	98.51	**99.35**	**96.39**	**97.68**	**98.89**	**95.10**	**97.40**	**96.20**
	Full-open	76.88	92.74	77.96	48.21	74.36	76.89	64.01	83.46	65.16
	Random	74.77	92.58	78.52	50.36	75.42	82.90	66.05	85.27	69.24
Learn	LwoC	92.36	**99.17**	95.77	**87.67**	**95.00**	88.92	89.19	**95.00**	**92.75**
	LwC-Sim	93.12	99.02	91.11	86.54	94.18	87.11	89.31	94.23	90.89
	LwC-TV	**93.23**	**99.04**	**96.18**	86.37	94.16	**91.63**	**89.77**	94.10	90.38
	LwC-Inv	85.78	92.38	84.54	75.60	93.49	73.44	84.51	91.47	83.23
	LwC-RIP	91.12	95.18	93.20	78.62	89.53	81.31	88.80	92.67	86.68

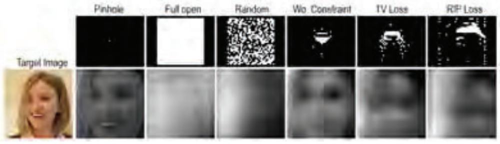
Best results are in red bold, second best in **bold**.

Trade-off between Human and Machine Vision



Real implementation





大阪大学
OSAKA UNIVERSITY

ECCV
TEL AVIV 2022

Privacy Preserving Action Recognition via Motion Difference Quantization

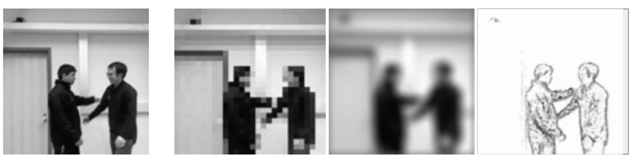

Sudhakar Kumawat Hajime Nagahara

Introduction

Hugging

Introduction

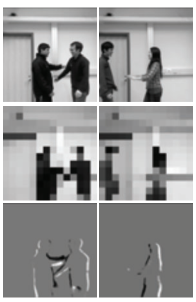
Some Common methods for privacy-preserving action recognition.



Original Scene Downsampled Blurred DVS Sensor

Goal: To develop an efficient encoder for the camera system that allows important features for action recognition while protecting actor(s) visual privacy.

Why privacy-preserving action recognition is hard?

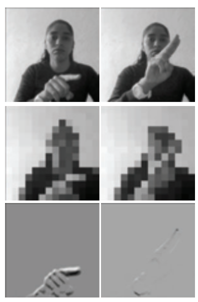


Action Recognition accuracy depends on:

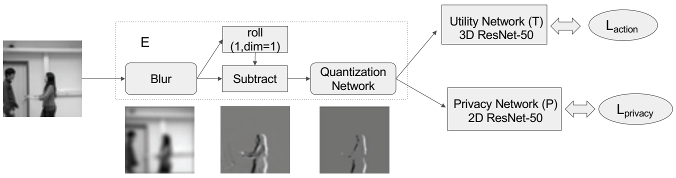
1. Spatial information
2. Temporal information

If the resolution of either of these information drops for protecting privacy, the action recognition accuracy also drops.

BDQ



Blur Difference Quantization



Given a frame v_i, we define a video as $V = \{v_i | i = 1, 2, ..., t\}$ where t is the number of frames.

Blur module: $B_{v_i} = G_\sigma v_i$, where $G_\sigma = \frac{1}{2\pi\sigma^2}exp(-\frac{x^2+y^2}{2\sigma^2})$

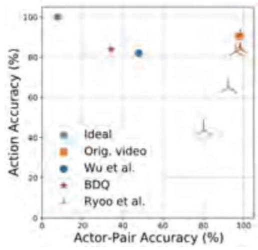
Difference module: $D(B_{v_i}, B_{v_j}) = B_{v_i} - B_{v_j}$

Quantization module: $Q(D(B_{v_i}, B_{v_j})) = \sum_{n=1}^{N-1} \sigma(H(D(B_{v_i}, B_{v_j}) - b_i))$, where $N = 16$, t: Temperature, $b_i = \{0.5, 1.5, \ldots, N-1.5\}$

BDQ Training: Repeat following two steps iteratively until convergence.

1. Fix P, train E and T using loss function $\mathcal{L}(V, \theta_E, \theta_T) = \mathcal{XE}(T(E(V)), L_{action}) - \alpha\mathcal{E}(P(E(V)))$
2. Fix E and T, train P using loss function $\mathcal{L}(V, \theta_P) = \mathcal{XE}(P(E(V)), L_{privacy})$

Evaluation on SBU dataset (Actions-8, Privacy-13)



Method	Params.	Size	FLOPs
Wu *et al*	1.3M	3.8Mb	166.4G
BDQ	16	3.4Kb	120.4M

Wu et al. "Privacy-preserving deep action recognition: An adversarial learning framework and a new dataset." IEEE Transactions on Pattern Analysis and Machine Intelligence (2020).

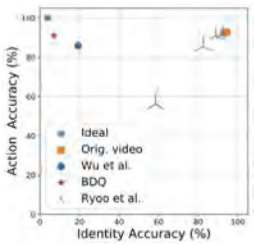
The above paper that we compare with use a UNet like encoder-decoder for video degradation.

Ryoo et al. "Privacy-preserving human activity recognition from extreme low resolution." AAAI (2017)

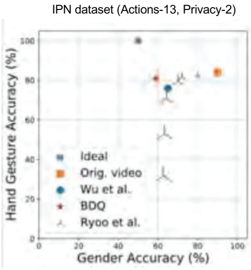
The above paper use downsampling for video degradation.

Evaluation on KTH and IPN datasets

KTH dataset (Actions-6, Privacy-25)

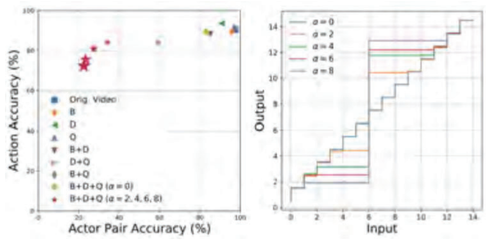


IPN dataset (Actions-13, Privacy-2)

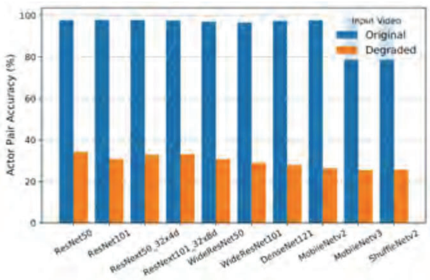


Ablation Study

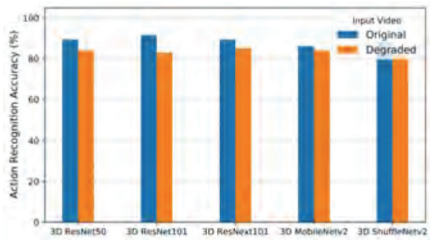
Here, we study the contribution of each component of BDQ in preserving privacy.



Strong Privacy Protection Features

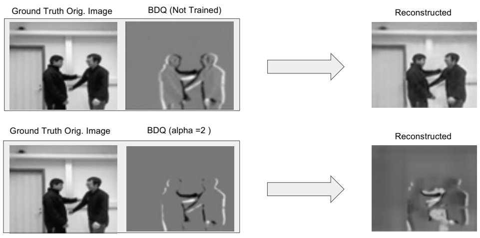


Generalized Action Recognition Features



Reconstruction Attack

Visualization of reconstruction by a 3D UNet model (attacker) trained with input as BDQ output videos and labels as original videos.



Conclusions

- Camera is a convenient equipment which can obtain a detailed scene information.
- However, it also obtains unwanted visual privacy.
- We proposed to jointly optimize the hardware, optics and sensor, and software, classification model.
- It realize the good balance of the privacy and task performances.