Addressing Privacy Concerns in Depth Sensors

Wiktor Mucha[0000−0002−6048−3425] and Martin Kampel[0000−0002−5217−2854]

Computer Vision Lab, TU Wien, Vienna, Austria
{wiktor.mucha,martin.kampel}@tuwien.ac.at

Abstract. Image-based assistive solutions raise concerns about the privacy of the individuals being monitored. The issue involves the situation when such technology is used in medical institutions to protect patients' health and support the personnel. These devices are installed in facilities and process images that include personal and behavioral data during the day. Other types of images than RGB are used to maintain privacy in this type of application, like depth images. Usage of depth cameras in the majority of publications is considered private protective. This paper discusses the issue of privacy in vision-based applications using depth modality. The factors affecting privacy in depth images are presented. The main problem that makes an image non-private is that the subjects' faces allow identification. This paper compares the Face Recognition (FR) technique between RGB and depth images. In the experimental part, a state-of-the-art model for FR in depth images is developed, which is used to establish boundary conditions when a person is recognized. The performance of FR between these two modalities is compared on two existing datasets containing images in both versions, including the training process. The study aims to determine under which conditions depth cameras preserve privacy and how much privacy they reveal.

Keywords: Privacy· Depth sensors· Multimodal vision· Face recognition· AAL

1 Introduction

Camera-based systems are used in assistive devices such as Active Assisted Living (AAL) solutions. They are installed in all kinds of medical facilities and by observing patients, reveal their private data which is questionable and may be unacceptable. It creates a demand for privacy-preserving solutions. A recurring method to face this is to replace the RGB camera with a depth sensor. This work is motivated by conflicting assumptions about privacy in depth-based applications. Some papers consider depth imaging to be privacy-preserving[2], and some do not[7]. Biometric features available in depth image are studied to determine how private it is. Various situations of revealing privacy are discussed. The main focus is put on Face Recognition (FR) from depth data which allows automatic recognition of subjects. In related work, the FR task is dominated by Deep Learning (DL) algorithms that replace traditional feature extraction methods. Their accuracy reaches almost 100% and depends on the image resolution. This

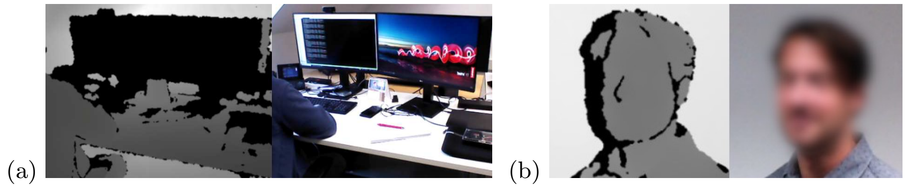

Fig. 1. Example of same scenes visible in RGB and depth modality.

study answers whether depth camera-based systems preserve subjects' privacy, under what circumstances, and what are the performance differences between RGB and depth FR. The experimental part aims to find the border conditions of correct identification by seeking minimal face resolution that allows this process. Further, it compares FR performance between these two modalities. The depth FR is less robust than RGB, and the learning process is slower. However, identification from depth data is correct under specific factors, showing that this modality does not fully preserve a person's identity. FR is correct even when faces are downsized to 40% of the original dimensions.

This paper is organized as follows. Section 2 describes privacy concerns in depth images, conditions for subjects recognition, facial features which permit FR in depth images, legal aspects of FR and personal privacy, and related work in FR from depth data. Section 3 introduces our experiment with results. Section 4 highlights key parts of this paper and its message.

2 Privacy in depth images

There is no single definition of privacy existing. This work takes deeper dive into this topic from a legal aspect. We consider data private when it does not reveal a person's identity or sensitive information. This statement is expanded more in Section 2.3. In contrast to RGB-based techniques in which subjects are monitored, depth sensors have an opinion of being privacy-protecting devices [2]. In this publication, the authors choose depth images to create a privacy-preserving fall detection system, and the sensor's privacy is taken for granted. Fall detection system [22] considers depth data as private but to additionally ensure, the system does not store any depth data during monitoring. On the contrary, Chou et al.[7] who design an Activity Recognition (AR) system for healthcare usage, address depth sensors as not privacy-preserving due to existing works performing FR form depth images. They employ their system for hand hygiene analysis where sensor captures faces. The image resolution is downsized to remove privacy-relevant information, and the activity recognition utility is retained. There are more such contradictory examples of publications.

Depth sensors do not preserve texture information, which protects against disclosing information related to colors. Only shapes are preserved in this type of image. Fig. 1(a) depict same scene in RGB and depth modality. The lack of available colors hides what is visible on a computer screen. In a real-life scenario, depth image protects the AAL device users' privacy, e.g., by not revealing information about what people are watching on TV. Certain shows or channels indicate their preferences, which they may consider private.

A distinguish problem is an acceptance of video-based AAL systems. Banerjee et al. [4] acknowledge this problem and present depth imaging technologies to their patients with an explanation. During the test of fall detection application in a hospital environment, depth data is presented to hospital staff and patients who consider it more private and acceptable than RGB. Ballester et al.[3] introduce a toileting assistance system for people with dementia. The concept is discussed with healthcare professionals from a facility for dementia patients. The outcome shows acceptance and trust for privacy in depth sensors, even in such an intimate environment as a toilet. The monitoring with depth image is fully accepted by 8 out of 13 people, where all participants reject RGB based system.

2.1 Recognition scenarios in depth images

Privacy of depth data differs between circumstances, and each application should be judged differently based on the environment. In health facilities, the personal identity is revealed from a human perspective when people vary in geometry. In the case of a group of 30 people, one of them is recognizable when it is a disabled person in a wheelchair, is taller or shorter than the others, or is using a walking support device like a walking frame. Viewing a situation when there are only images or videos available without any additional data, what allows automatically recognize people are biometric features, e.g., face, height, gait, skin color, tattoos, and scars [19]. Depth images do not contain the last three of them, enhancing privacy by reducing the number of information allowing identification. In the literature, there are publications presenting gait recognition. Dubois et al.[10] use depth images to extract gait pattern using Hidden Markov Model (HMM) to recognize people inside the house. In their experiment, they detect correctly 17 individuals out of 20. However, such tests are too narrow to compare with real-world conditions. Publications referenced in Section 2.4 show FR allows correct identification in available depth datasets. However, most test data has a low number of subjects (around 20). Complete FR pipeline for identification from a depth camera includes face detection at an early stage. These detectors tend to perform with lower accuracy than RGB, directly impacting FR systems' performance which favors RGB-based models at the very first step[18].

2.2 Face features for depth FR

FR is based on facial features which DL models extract. Abudarham et al. [1] shows that the most impactful features for FR are landmarks, e.g., eye shape, eyebrows, and lips. They hypothesize that Convolutional Neural Networks (CNN) employ similar features for FR as human beings. Looking at the depth image, there are fewer features visible from a human perspective than in the RGB modality. However, their visibility is linked with the resolution of the sensor. What can be seen by humans are basic shapes of hair, nose, lips, or eyes, but it lacks texture information (Fig. 1(b)). In studies of Kendrick et al.[15] authors detect shapes of mouth, nose tip, nostrils, eyes, and forehead using DL networks in depth images. Nevertheless, depth modality performs worse than RGB by achieving the largest error in their position estimation, and models are tested in laboratory conditions, which does not answer how accurate it works in the wild.

2.3 Legal aspects of privacy and FR

The concept of privacy is not explicitly normalized. The right to it is centrally standardized in Art. 8 (1) European Convention on Human Rights (ECHR), according to which everyone has the right to respect for his private and family life, his home, and his correspondence. Further standards are in Art. 7 of the Charter of Fundamental Rights of the European Union (CFR), which regulates violations of privacy and certain professional secrets. The right to data protection is also understood as a facet of privacy. Art. 8 of the CFR standardizes the right to the protection of personal data for every person. Art. 1 of the General Data Protection Regulation (GDPR) specifies the protection when personal data is processed. The GDPR can be understood as a substantive concretization of Art. 8 of the CFR and as a formulation of it in Union law.

Except reviling personal information, identity identification invades privacy. Regarding the study[14], forensic FR becomes a ubiquitous tool to guide court investigations. Automatic FR systems lack methodological standardization and empirical validation, placing FR below fingerprints or DNA in terms of trust. In continental Europe, there is no legal limit on the admissibility of scientific evidence. Judges evaluate the relevance of the evidence on their own according to the state of available scientific knowledge. In the USA, according to the Federal Rule of Evidence 702, the evidence has to be accepted by the scientific community, and the employed method has to have known error rates and be available for peer review and publication.

2.4 Face Recognition in depth images

Face recognition is the task of identifying people by their faces. First, the face has to be detected and cropped. Further image is processed to extract features describing the face. RGB is not employed in depth recognition, and features are extracted from 3d data. In the literature, there are works performing FR from depth data using feature extractors like Soon-kak Kwon[16] who starts the framework with a face detector based on depth data only. Further, features are extracted by applying the 3D-LBP descriptor. DL models overtake traditional descriptors. Feng et al.[12] presents a system with an additional DL network for normalizing facial images into the frontal pose and neutral expression, reducing noise and reconstructing missing regions. A second CNN performs feature extraction and recognition task. Another DL approach is presented by Hu et al.[13] who boost recognition from low-quality data employing high-quality samples. This method is restricted by the low availability of datasets, including low and high-resolution images. In [21] normalization impact is studied for FR task.

Table 1. Comparison of FR top results on depth datasets, which differ in image quality and a number of subjects.

Databases	No. of Subjects	Sensor	Resolution	Sensor Precision [mm]	Accuracy [%]
BU-3DFE	100	3dMD	8K	0.2	99.3 [20]
Lock3DFace	509	Kinect II	20K	$\geq$2	86 [12]
IIIT-D	106	Kinect I	13K	2-4	26.8[8]
CurtinFaces	52	Kinect I	13K	2-4	89.9[5]

Models are validated with different variants of pre-processing. According to this study, these techniques combined with CNN decrease recognition accuracy because applying them reduces the 3D content available for the model.

Table 1 lists examples of datasets with a different number of subjects, captured with various sensors distinct in resolution and precision. Higher accuracy is achieved in works using higher precision sensors and datasets with fewer subjects, which does not answer how accurate FR is in the wild.

3 The experiment

In this study, the DL model[23] is implemented and trained for depth FR to achieve state-of-the-art performance. Such architectures are proved by literature in Section 2.4 to perform robustly on depth images for FR. The learning process is performed on combined *BIWI*[11], *Pandora*[6], and *UMBDB*[9] datasets. They all include depth images of people's faces taken in laboratory conditions. The resulting model of our process is compared with existing works by tests on a separate set of data. The test subset of *HRRFaceD Database*[17] is taken to perform cross-dataset validation. Every image in test data has a predicted vector of embeddings. From predictions, image pairs are composed, and the distance is calculated between their vectors. The predicted label of input images is the class of the image with which the minimum pair is obtained. Our model is placing second in accuracy with 97.95% after Borghi et al.[5] who achieve 98.90%.

In real-world conditions, the correctness of FR depends on the distance between the subject and the sensor, which comes down to the face size in pixels. When the subject's face is near, its size is close to sensor maximum resolution. With the distance increasing, the face dimension reduces. A larger resolution of the face provides higher quality descriptions of facial landmarks, which results in better performance. Analysis of Table 1 confirms this thesis. To find the border conditions of correct identification, minimal resolution of the face allowing this process is searched. Faces from *HRRFaceD* are scaled-down, and the FR accuracy is calculated for each subset created.

A decrease occurs when the face is smaller than 40% of the original size (Fig. 2(a)). Since the test faces vary in dimensions, their average values are computed. The original size is 118x153 pixels, and for 40% it is 47x61 pixels. The worst FR results are seen below 10% (11x14 pixels).

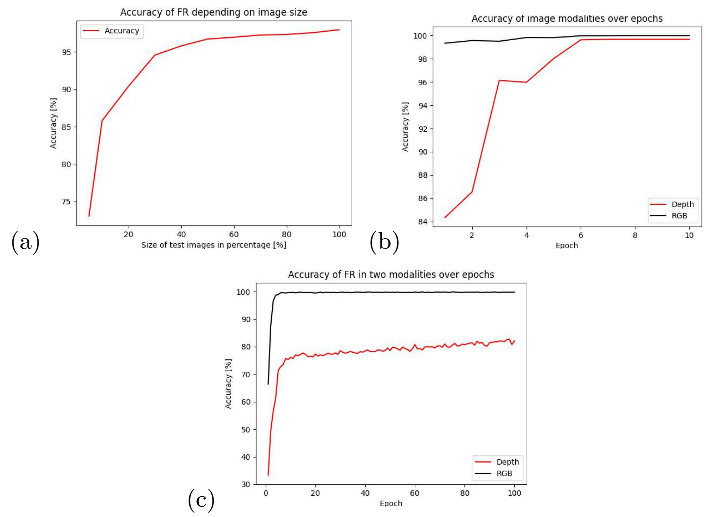

Fig. 2. Accuracy of depth FR depending on image size (a) and on epochs with RGB and depth FR trained form scratch on *Pandora*(b) and *BIWI*(c).

The second part of our experiment compares RGB and depth modalities. Firstly, the best possible models are viewed. The training is performed on *Pandora*, *UMBDB* and *HRRFaceD* and test data is randomly chosen 2000 images from *BIWI* where every depth image has a corresponding RGB frame. For RGB modality *vggface2* model is fine-tuned on *Pandora*. The accuracy is calculated five times, every iteration on a different subset, and the average value is given to reduce a random factor. The result is 98.79% accuracy for RGB and 80.50% for depth modality. A similar process is performed with tests on *Pandora*. This scenario gives a better overview of potential FR accuracy in real-world conditions due to the various occlusions in images, where previously tested *BIWI* does not include any. Learning is done on *UMBDB*, *HRRFaceD* and *BIWI*. RGB performs with 85.15% accuracy and depth scores 53.23%. Secondly, both models are trained from scratch on the same data with similar parameters. The experiment is carried out twice. *Pandora* used for training and *BIWI* as a test results in 98.39% (RGB) and 79.38% (depth). Replacement of subsets with tests on *Pandora* gives 84.60% (RGB) and 48.74% (depth) accuracy. For depth images, the learning process is observed to be longer (Fig. 2(b)). When trained on *BIWI*, the validation of RGB reaches ≈100% before the 10th epoch. In depth modality, 100 epochs are not enough to get close to RGB performance (Fig. 2(c)).

4 Conclusion

Under the law, everyone has a right to privacy. The available publications are not consistent with the privacy of depth images which is problematic with respect to legal regulations. The main advantage of them is the protection of data that can be considered personal by subjects of visual monitoring and are visible in RGB images. Regarding the identity reveal, FR in depth is based on similar face features like in RGB. A comparison between these modalities favored RGB over depth with a significant accuracy difference. An examination of the learning process of our model with its performance confirmed more challenging feature extraction for a depth study. Whatmore, the high accuracy in depth FR was achieved in datasets with a low number of subjects that were not comparable with conditions in health facilities and real-world scenarios with additional environmental distortions, a decrease of depth face detection accuracy, and sensor's range. On the other hand, DL methods showed improvement in the accuracy of FR, even on small face resolutions. These circumstances make depth data more private than RGB, but when designing depth-based monitoring solutions, it is necessary to consider the possible disclosure of subjects' identities, when the number of individuals in the facility is small (less than 100), data is gathered with high precision sensors. Privacy preservation is also determined by the environment in which the system is used and to whom it is addressed, what activities are monitored, and what data can be recognized.

Acknowledgements This work was partly supported by VisuAAL ITN H2020 (grant agreement No. 861091) and the Austrian Research Promotion Agency (FFG) under the Grant Agreement No. 879744.

References

1. Abudarham, N., Shkiller, L., Yovel, G.: Critical Features for Face Recognition. vol. 182, pp. 73–83. Elsevier (2019). https://doi.org/10.1016/j.cognition.2018.09.002
2. Arulselvi, G., Poornima, D., Anand, S.J.: Privacy Preserving Elderly Fall Detection Using Kinect Depth Images Based on Deep Convolutional Neural Networks. European Journal of Molecular & Clinical Medicine **7**(3), 5492–5510 (2020)
3. Ballester, I., Mujirishvili, T., Kampel, M.: RITA: A Privacy-Aware Toileting Assistance Designed for People with Dementia. In: Pervasive Computing Technologies for Healthcare. pp. 318–330. Springer International Publishing (2022). https://doi.org/10.1007/978-3-030-99194-4_20
4. Banerjee, T., Rantz, M., Li, M., Popescu, M., Stone, E., Skubic, M., Scott, S.: Monitoring Hospital Rooms for Safety Using Depth Images. AI for Gerontechnology, Arlington, Virginia, US (2012)
5. Borghi, G., Pini, S., Vezzani, R., Cucchiara, R.: Driver Face Verification With Depth Maps. Sensors **19**(15), 3361 (2019). https://doi.org/10.3390/s19153361
6. Borghi, G., Venturelli, M., Vezzani, R., Cucchiara, R.: Poseidon: Face-From-Depth for Driver Pose Estimation. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 5494–5503. IEEE (2017). https://doi.org/10.1109/CVPR.2017.583
7. Chou, E., Tan, M., Zou, C., Guo, M., Haque, A., Milstein, A., Fei-Fei, L.: Privacy-Preserving Action Recognition for Smart Hospitals Using Low-Resolution Depth Images. arXiv preprint arXiv:1811.09950 (2018)

8. Chowdhury, A., Ghosh, S., Singh, R., Vatsa, M.: RGB-D Face Recognition via Learning-Based Reconstruction. In: 2016 IEEE 8th International Conference on Biometrics Theory, Applications and Systems (BTAS). pp. 1–7. IEEE (2016). https://doi.org/10.1109/BTAS.2016.7791199
9. Colombo, A., Cusano, C., Schettini, R.: UMB-DB: A Database of Partially Occluded 3D Faces. In: 2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops). pp. 2113–2119. IEEE (2011). https://doi.org/10.1109/ICCVW.2011.6130509
10. Dubois, A., Bresciani, J.P.: Person Identification From Gait Analysis With a Depth Camera at Home. In: 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). pp. 4999–5002. IEEE (2015). https://doi.org/10.1109/EMBC.2015.7319514
11. Fanelli, G., Dantone, M., Gall, J., Fossati, A., Van Gool, L.: Random Forests for Real Time 3D Face Analysis. International Journal of Computer Vision **101**(3), 437–458 (2013). https://doi.org/10.1007/s11263-012-0549-0
12. Feng, Z., Zhao, Q.: Robust Face Recognition with Deeply Normalized Depth Images. In: Chinese Conference on Biometric Recognition. pp. 418–427. Springer (2018). https://doi.org/10.1007/978-3-319-97909-0_45
13. Hu, Z., Gui, P., Feng, Z., Zhao, Q., Fu, K., Liu, F., Liu, Z.: Boosting Depth-Based Face Recognition from a Quality Perspective. Sensors **19**(19), 4124 (2019). https://doi.org/10.3390/s19194124
14. Jacquet, M., Champod, C.: Automated Face Recognition in Forensic Science: Review and Perspectives. Forensic Science International **307**, 110124 (2020). https://doi.org/10.1016/j.forsciint.2019.110124
15. Kendrick, C., Tan, K., Walker, K., Yap, M.H.: Towards Real-Time Facial Landmark Detection in Depth Data Using Auxiliary Information. Symmetry **10**(6), 230 (2018). https://doi.org/10.3390/sym10060230
16. Kwon, S.k.: Face Recognition Using Depth and Infrared Pictures. Nonlinear Theory and Its Applications, IEICE **10**, 2–15 (01 2019). https://doi.org/10.1587/nolta.10.2
17. Mantecón, T., del Blanco, C.R., Jaureguizar, F., García, N.: Visual Face Recognition Using Bag of Dense Derivative Depth Patterns. IEEE Signal Processing Letters **23**(6), 771–775 (2016). https://doi.org/10.1109/LSP.2016.2553784
18. Mucha, W., Kampel, M.: Depth and Thermal Images in Face Detection - A Detailed Comparison Between Image Modalities. In: 2022 the 5th International Conference on Machine Vision and Applications (ICMVA) (ICMVA 2022). ACM (2022). https://doi.org/10.1145/3523111.3523114.
19. North-Samardzic, A.: Biometric technology and ethics: Beyond security applications. Journal of business ethics **167**(3), 433–450 (2020)
20. Ocegueda, O., Passalis, G., Theoharis, T., Shah, S.K., Kakadiaris, I.A.: UR3D-C: Linear Dimensionality Reduction for Efficient 3D Face Recognition. In: 2011 International Joint Conference on Biometrics (IJCB). pp. 1–6. IEEE (2011). https://doi.org/10.1109/IJCB.2011.6117521
21. Pini, S., Borghi, G., Vezzani, R., Maltoni, D., Cucchiara, R.: A Systematic Comparison of Depth Map Representations for Face Recognition. Sensors **21**(3), 944 (2021). https://doi.org/10.3390/s21030944
22. Planinc, R., Kampel, M.: Introducing the Use of Depth Data for Fall Detection. Personal Ubiquitous Comput. **17**(6), 1063–1072 (aug 2013). https://doi.org/10.1007/s00779-012-0552-z
23. Schroff, F., Kalenichenko, D., Philbin, J.: Facenet: A unified embedding for face recognition and clustering. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 815–823 (2015)