Beyond Privacy of Depth Sensors in Active and Assisted Living Devices

Wiktor Mucha
wiktor.mucha@tuwien.ac.at
TU Wien, Computer Vision Lab
Vienna, Austria

Martin Kampel
martin.kampel@tuwien.ac.at
TU Wien, Computer Vision Lab
Vienna, Austria

ABSTRACT

Active and Assisted Leaving (AAL) devices that use cameras and images raise concerns about the privacy of the monitored individuals. These devices capture images that include personal and behavioral data during the day. Most authors decide to switch from RGB to depth sensors to maintain privacy. Nevertheless, not all available works agree that depth image is private, which creates an open legal problem for AAL applications. In this paper, privacy is discussed in vision-based systems using depth sensors. Various factors of depth and RGB images that might affect privacy are presented to define the privacy level of depth devices. One of the main issues that make an image non-private is that the subjects' faces are visible and can be identified. In the experimental part, a state-of-the-art Face Recognition (FR) model in depth images is developed. It is used to establish boundary conditions allowing correct recognition of a person's face. A comparison between FR in RGB and depth images is performed, including the ability to learn the model by training the two modalities from scratch on identical data. This study answers under which conditions depth cameras protect the privacy and how much privacy is disclosed by them.

CCS CONCEPTS

• **Computing methodologies** → *3D imaging*.

KEYWORDS

image modalities, privacy, depth sensors, face recognition, AAL

ACM Reference Format:
Wiktor Mucha and Martin Kampel. 2022. Beyond Privacy of Depth Sensors in Active and Assisted Living Devices. In *The15th International Conference on PErvasive Technologies Related to Assistive Environments (PETRA '22), June 29-July 1, 2022, Corfu, Greece.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3529190.3534764

1 INTRODUCTION

Systems based on cameras in the field of Active and Assisted Living (AAL) bring controversies with privacy. These devices are used in hospitals to monitor patients, so the data they collect may be considered private and sensitive. Privacy is a subjective topic depending on every individual. In this work, it is understood as a protection of their identity or information that may be unwanted to reveal by them. Researchers in the AAL field investigate the possibility of switching RGB sensors to depth cameras to increase privacy levels and define this image modality as privacy-preserving[2, 17, 30]. Despite this, there are articles that claim that depth images reveal individual privacy, such as [6, 18]. The contrary opinions and unclear status of privacy of depth sensors motivate this work and create a potential legal problem for these systems. This article looks at privacy-preserving AAL applications and discusses available biometric features in depth images and necessary scenarios for privacy enhancement. Our main focus is on Face Recognition (FR) systems, which are most successful in identifying subjects from depth data. FR algorithms mainly employ Deep Learning (DL) architectures and reach results nearly identical to 100% (See Table 1). This work answers the questions about the privacy of depth sensors and needed circumstances for its preservation and differences in automatic recognition using subjects' faces between RGB and depth modality, including performance and learning process. In the experimental part, state-of-the-art FR architecture is implemented. In depth data, recognition is less accurate than in RGB images, and training takes longer. Nevertheless, in available datasets, identifications from depth images are correct, showing their potential to reveal biometric features.

This paper is organized as follows. Section 2 defines privacy and its legal aspects. Privacy concerns in depth images and motivation of this work are discussed in Section 2.1. Section 2.2 presents available biometrics in images and circumstances for recognition, including automatic and manual personal identification. Section 2.3 explains facial features which permit accurate FR in depth images. Section 2.4 presents related work in FR from depth data. Section 3 introduces our experiment with results and discussion. Section 4 concludes this work by highlighting key parts of this paper and its message.

PETRA '22, June 29-July 1, 2022, Corfu, Greece

ACM ISBN 978-1-4503-9631-8/22/06…$15.00
https://doi.org/10.1145/3529190.3534764

2 PRIVACY IN DEPTH IMAGES

Personal privacy is a subjective matter and depends on a variety of factors, such as cultural background and origin[15]. The Art. 8 (1) of the European Convention on Human Rights (ECHR) and Art. 7 of the Charter of Fundamental Rights of the European Union (CFR) standardize the right to privacy for every individual, their family lives, homes, correspondence, and its violations. The term privacy also refers to the protection of personal data as described in Art. 8 of the CFR and Art. 1 of the General Data Protection Regulation (GDPR), which is oriented toward personal data processing.

Table 1: Comparison of top results on depth data only in FR, different data sets

Databases	No. of Subjects	Sensor	Resolution	Sensor Precision [mm]	Accuracy [%]
FRGC v2	466	Vivid 910	60K	0.1	99.6 [28]
BU-3DFE	100	3dMD	8K	0.2	99.3 [24]
Lock3DFace	509	Kinect II	20K	≥2	86 [11]
RGBD-W	2239	RealSense	45K	≥2	64[13]
IIIT-D	106	Kinect I	13K	2-4	26.8[7]
CurtinFaces	52	Kinect I	13K	2-4	89.9[3]
Eurecom	52	Kinect I	13K	2-4	69.7[22]

Fig. 1: Example of the same scene visible in RGB and depth modality.

2.1 View at privacy in depth sensors

Depth sensors are considered privacy preserving by various authors who design fall detection or action recognition AAL systems [2, 5, 17, 30]. All of them state the privacy of such sensors as granted, unquestionable, and indisputable. Planinc and Kampel[27] present a fall-detection algorithm, and they highlight the depth image as similar to an anonymized RGB frame. On the other hand, Park et al.[25] with work about sign language translation, describes depth images as less intrusive to privacy in comparison to RGB. Kong et al.[18] implement a privacy-preserving fall detection system considering depth image as not fully private. Their system is based on a depth sensor like other publications referenced in this section. Since depth data is considered not private by them, they encode input video streams with fast Fourier Transform for additional security in case of a potential data leak. A similar belief share Chou et al.[6] who design an activity recognition system for healthcare usage. They address the problem of privacy preservation in depth images by downsampling resolution. Their post-processed image reveals as little information for potential identification as possible while performing activity recognition.

Regarding personal data revealed with camera monitoring, the depth sensor, contrary to RGB cameras, does not preserve texture information. Lack of texture hides potentially unwanted information in captured images, e.g., what a person is wearing, reading, or watching. This increases privacy level even when personal identity is revealed, but it is not possible to correlate sensitive data with a recognized person, e.g., what a subject is watching on TV (See Fig. 1).

2.2 Recognition scenarios in depth images

Recognition of subjects is possible by reading their biometric features. Using camera-based systems, features that can be analyzed are a face, height, scars, and gait[29]. Depth sensors do not prevail texture information, which hides skin color data, scars, and tattoos. During the design of the AAL system with privacy preservation, it is essential to define circumstances on the environment and what personal information is revealed. In small facilities, for instance with a number of patients restricted to 30, persons with unique characteristics, e.g., significantly taller or smaller than others or those who use a wheelchair or a walking frame, are identifiable. From an algorithmic point of view using depth images, some authors propose gait recognition systems like Dubois et al.[9]. Nevertheless, the presented tests are too narrow to compare with a real-world scenario. FR shows more promising results, which is discussed in Section 2.4 where in tested datasets, the accuracy of systems is reaching close to 100%. The conclusion coming from the analysis of best results in available datasets (Table 1) is that FR performance relies on sensor precision, resolution, and a number of subjects. In the wild, the first two translate to the face size range, which makes it identifiable.

Jacquet et al. [14] point out forensic FR becomes a universal tool to guide court investigations. These systems currently lack methodological standardization and empirical validation, making them less reliable than fingerprints or DNA. From a legal aspect, such technologies might be employed by courts. There is no legal limit on the admissibility of scientific evidence in continental Europe. Judges assess the applicability of the proof on their own according to the state of available scientific knowledge. In the USA, according to the Federal Rule of Evidence 702, proof has to be accepted by the scientific community, and the employed technique has to have understood error rates and be available for peer review and publication.

2.3 Face features for depth FR

Humans recognize people's faces by their high-level features like eyes, shapes, or skin tone. With DL algorithms, it is not clear what data is exactly extracted for predictions, but regarding the study of Abudarham et al. [1] models are using similar landmarks as we humans do. Authors point out eye-shape, eyebrows, and lips as the most impactful features for DL networks performing FR. Depth image contains fewer features visible to our eyes because of missing texture data. However, basic shapes of hair, nose, lips, or eyes are noticeable (See Fig. 2). Studies in the field of facial

Fig. 2: Face example in depth and RGB modality.

Fig. 3: Examples of occluded faces from *Pandora* dataset.

landmark detection from depth data of Kendrick et al.[16] show the availability of DL networks to detect shapes of mouth, nose tip, nostrils, eyes, and forehead. Nevertheless, depth modality performs worse than RGB by achieving the largest error in their position estimation. Further uncertainty is in test data employed in this study because it is captured in laboratory conditions.

2.4 Face recognition in depth images

One of the first FR methods based on depth images is the work of Rui Min et al.[21]. However, the solution presented in this paper is using Viola-Jones face detection exploited on RGB inputs. Detected faces are used for cropping the depth image. Point cloud from each face region is obtained. Further, the surface is segmented into smaller regions, each of them having an assigned weight. Regions that are potentially unstable features, e.g., hair are filtered out. In the conducted experiment with faces of 20 people, the accuracy is 97.7%.

FR from only depth data is presented by Feng et al.[11]. They propose a novel method to normalize facial images to frontal pose and neutral expression for recognition enhancement. The framework consists of two Convolutional Neural Networks (CNNs). The first one converts and reconstructs 3d image as a 3DMM model. The generated model is normalized to reduce noise and reconstruct missing regions. Further, a second CNN performs feature extraction and recognition task. The mean accuracy is 86% on the Lock3DFace database, and it is a top score during the release of this paper.

A comparison between performing FR from depth and infrared (IR) data is presented in the work of Soon-kak Kwon[19]. In this work, the author proposes a fusion of depth and IR images. The process starts with a face detector based on depth data only. The algorithm locates the nose position from the image and subtracts the background. From the separated image of the face, features are extracted. It is done by applying the 3D-LBP descriptor. The first four layers are used to obtain histograms describing features. Later those histograms are used for an identification process. In the experiment, there are 20 participants, and each of them is pictured from different perspectives and light conditions. With IR images, recognition accuracy reaches 98%, and with the depth data, it is 93% due to the noises. The constraints are the accuracy drop when faces are tilted and poor face detection depending on nose position.

Authors[12] are boosting recognition form low-quality data using high-quality samples. There are three different techniques proposed where the model is enhanced during training by higher resolution images. The main obstacle is the lack of available data sets, including both low and high-resolution images.

In [26] normalization impact is studied for FR task. Models like VGG or ResNet are validated with different variants of preprocessing like filtering, hole filling, and equalization. The summary is that all of those techniques in combination with CNNs decrease overall accuracy because applying them reduces the 3D content available for the model. Such models are also less capable of generalizing and transferring to other conditions.

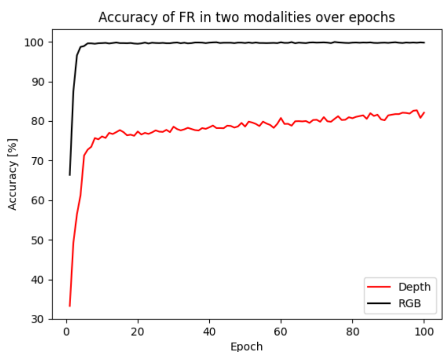
3 THE EXPERIMENT

A state-of-the-art DL model is implemented and trained for depth FR in this study. In the literature, such architectures are proven to perform well for FR in depth images. The learning process is performed on combined *BIWI*[10], *Pandora*[4], and *UMBDB*[8] datasets. In order to compare the result of our process with other existing works, it is tested with external data. In order to perform cross-dataset validation, a subset of *HRRFaceD Database*[20] is used as a benchmark. There is predicted an embedding vector for every image in test data. Using the predictions, image pairs are derived, and their vector distances are calculated. Predicted labels are associated with the image class of the pair with minimum distance obtained. The accuracy of our model is 97.95%, just behind Borghi et al.[3], who achieve 98.90%.

3.1 Recognition performance in relation to the face size

FR's correctness in real-world conditions depends on several factors. The pipeline starts with a face detector which in depth images tend to perform less accurately than in RGB image[23] and is a first constrain. Secondly, the distance from the sensor to the subject and the sensor's resolution is substantial, which transfers to the face size in pixels. The objective of this study is to identify the minimal resolution of the face that allows correct identification. Accuracy is calculated for multiple scales of original images from *HRRFaceD*. The results are presented in Fig. 4(a) with accuracy on the vertical axis and percentage of original image size on the horizontal axis. A significant drop appears for faces smaller than 40% of the original dimensions. In *HRRFaceD* every sample differs in its resolution. Due to this, the mean values are given. The original size is 118x153 pixels, and for 40%, it is 47x61 pixels. The worst FR results are seen below 10% (11x14 pixels).

(a)



(b)

Fig. 4: Accuracy of depth FR depending on image size (a) and on epochs with RGB and depth FR (b).

3.2 Performance of FR in RGB and depth image modalities

Our research compares RGB and depth modalities in its second part. It is first determined how best available models perform. *Pandora*, *UMBDB* and *HRRFaceD* are employed as learning data and for test 2000 images are randomly chosen from *BIWI* where every depth image has a corresponding RGB frame. RGB FR uses *vggface2* model which is additionally fine-tuned on *Pandora* for better results. A random factor caused by choosing test images from a more extensive set is reduced by calculating the accuracy five times on a separate subset and then taking the average. The result is 98.79% accuracy for RGB and 80.50% for depth modality. A process is repeated with tests on *Pandora*. This scenario presents a better overview of probable FR accuracy in a real-world environment because images include diverse occlusions, contrary to earlier tested *BIWI* which does not contain any. For training *UMBDB*, *HRRFaceD* and *BIWI* are used. RGB performs with 85.15% accuracy and depth scores 53.23%.

To see the difference in training and performance in equal conditions, both models are trained from scratch on the same data with similar parameters. The learning process is performed using *BIWI*. Comparison is done with *Pandora* and results with 84.60% (RGB) and 48.74% (depth) accuracy. For depth images, the learning process is observed to be longer, and 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. 4(b)).

4 CONCLUSION

There is no consistent information about the privacy of depth images in existing publications, which creates a potential legal problem for AAL applications employing this type of camera. Regarding legal regulations, everyone has a right to privacy, private data, and data protection during its processing. Using depth sensors offers the advantage of protecting information that can be considered personal and visible in RGB images during monitoring processes. Concerning identity reveal, FR in depth is based on similar facial features to FR in RGB, and a comparison between these modalities has a significant advantage for RGB over depth. After examining its learning process, a deeper study revealed a more challenging feature extraction for our model. Moreover, a high depth FR accuracy was observed in datasets with low numbers of subjects. This is not comparable to situations within health facilities and in real-world scenarios when additional environmental factors, sensor range, or depth estimation errors are present. In contrast, present DL methods improved FR accuracy even for small faces resolutions. Because of these circumstances, depth data is more private than RGB. Yet, it is necessary to consider the possibility of revealing subjects' identities when designing solutions based on depth monitoring, especially when a small number of individuals is present (less than 100) and data is gathered with high-precision sensors. In addition, the environment in which the system is used and to whom it is addressed, what activities are monitored, and what data can be recognized determine privacy preservation.

ACKNOWLEDGMENTS

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] Naphtali Abudarham, Lior Shkiller, and Galit Yovel. 2019. Critical features for face recognition. *Cognition* 182 (2019), 73–83.
[2] G Arulselvi, D Poornima, and Sujatha Jamuna Anand. 2020. PRIVACY PRESERVING ELDERLY FALL DETECTION USING KINECT DEPTH IMAGES BASED ON DEEP CONVOLUTIONAL NEURAL NETWORKS. *European Journal of Molecular & Clinical Medicine* 7, 3 (2020), 5492–5510.
[3] Guido Borghi, Stefano Pini, Roberto Vezzani, and Rita Cucchiara. 2019. Driver face verification with depth maps. *Sensors* 19, 15 (2019), 3361.
[4] Guido Borghi, Marco Venturelli, Roberto Vezzani, and Rita Cucchiara. 2017. Poseidon: Face-from-depth for driver pose estimation. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 5494–5503.
[5] Ming-Ching Chang, Ting Yi, Kun Duan, Jiajia Luo, Peter Tu, Michael Priebe, Elena Wood, and Max Stachura. 2017. In-bed patient motion and pose analysis using depth videos for pressure ulcer prevention. In *2017 IEEE International Conference on Image Processing (ICIP)*. IEEE, 4118–4122.
[6] Edward Chou, Matthew Tan, Cherry Zou, Michelle Guo, Albert Haque, Arnold Milstein, and Li Fei-Fei. 2018. Privacy-preserving action recognition for smart hospitals using low-resolution depth images. *arXiv preprint arXiv:1811.09950* (2018).
[7] Anurag Chowdhury, Soumyadeep Ghosh, Richa Singh, and Mayank Vatsa. 2016. RGB-D face recognition via learning-based reconstruction. In *2016 IEEE 8th International Conference on Biometrics Theory, Applications and Systems (BTAS)*. IEEE, 1–7.
[8] Alessandro Colombo, Claudio Cusano, and Raimondo Schettini. 2011. UMB-DB: A database of partially occluded 3D faces. In *2011 IEEE international conference on computer vision workshops (ICCV workshops)*. IEEE, 2113–2119.
[9] Amandine Dubois and Jean-Pierre Bresciani. 2015. 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). IEEE, 4999–5002.

[10] Gabriele Fanelli, Matthias Dantone, Juergen Gall, Andrea Fossati, and Luc Van Gool. 2013. Random forests for real time 3d face analysis. *International journal of computer vision* 101, 3 (2013), 437–458.

[11] Ziqing Feng and Qijun Zhao. 2018. Robust Face Recognition with Deeply Normalized Depth Images. In *Chinese Conference on Biometric Recognition*. Springer, 418–427.

[12] Zhenguo Hu, Penghui Gui, Ziqing Feng, Qijun Zhao, Keren Fu, Feng Liu, and Zhengxi Liu. 2019. Boosting Depth-Based Face Recognition from a Quality Perspective. *Sensors* 19, 19 (2019), 4124.

[13] Zhenguo Hu, Qijun Zhao, and Feng Liu. 2019. Revisiting Depth-based Face Recognition from a Quality Perspective. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*. 0–0.

[14] Maëlig Jacquet and Christophe Champod. 2020. Automated face recognition in forensic science: Review and perspectives. *Forensic science international* 307 (2020), 110124.

[15] Naz Kaya and Margaret J. Weber. 2003. Cross-cultural differences in the perception of crowding and privacy regulation: American and Turkish students. *Journal of Environmental Psychology* 23, 3 (2003), 301–309. https://doi.org/10.1016/S0272-4944(02)00087-7

[16] Connah Kendrick, Kevin Tan, Kevin Walker, and Moi Hoon Yap. 2018. Towards real-time facial landmark detection in depth data using auxiliary information. *Symmetry* 10, 6 (2018), 230.

[17] Michal Kepski and Bogdan Kwolek. 2014. Fall detection using ceiling-mounted 3d depth camera. In *2014 International conference on computer vision theory and applications (VISAPP)*, Vol. 2. IEEE, 640–647.

[18] Xiangbo Kong, Zelin Meng, Lin Meng, and Hiroyuki Tomiyama. 2018. A privacy protected fall detection IoT system for elderly persons using depth camera. In *2018 International Conference on Advanced Mechatronic Systems (ICAMechS)*. IEEE, 31–35.

[19] Soon-kak Kwon. 2019. Face recognition using depth and infrared pictures. *Nonlinear Theory and Its Applications, IEICE* 10 (01 2019), 2–15. https://doi.org/10.1587/nolta.10.2

[20] Tomás Mantecón, Carlos R del Blanco, Fernando Jaureguizar, and Narciso García. 2016. Visual face recognition using bag of dense derivative depth patterns. *IEEE Signal Processing Letters* 23, 6 (2016), 771–775.

[21] Rui Min, Jongmoo Choi, Gérard Medioni, and Jean-Luc Dugelay. 2012. Real-time 3D face identification from a depth camera. In *Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012)*. IEEE, 1739–1742.

[22] Rui Min, Neslihan Kose, and Jean-Luc Dugelay. 2014. Kinectfacedb: A kinect database for face recognition. *IEEE Transactions on Systems, Man, and Cybernetics: Systems* 44, 11 (2014), 1534–1548.

[23] Wiktor Mucha and Martin Kampel. 2022. 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)* (Singapore, Singapore) *(ICMVA 2022)*. Association for Computing Machinery, New York, NY, USA, 16–21. https://doi.org/10.1145/3523111.3523114

[24] Omar Ocegueda, Georgios Passalis, Theoharis Theoharis, Shishir K Shah, and Ioannis A Kakadiaris. 2011. UR3D-C: Linear dimensionality reduction for efficient 3D face recognition. In *2011 International Joint Conference on Biometrics (IJCB)*. IEEE, 1–6.

[25] HyeonJung Park, Jong-Seok Lee, and JeongGil Ko. 2020. Achieving Real-Time Sign Language Translation Using a Smartphone's True Depth Images. In *2020 International Conference on COMmunication Systems & NETworkS (COMSNETS)*. IEEE, 622–625.

[26] Stefano Pini, Guido Borghi, Roberto Vezzani, Davide Maltoni, and Rita Cucchiara. 2021. A Systematic Comparison of Depth Map Representations for Face Recognition. *Sensors* 21, 3 (2021), 944.

[27] Rainer Planinc and Martin Kampel. 2013. Introducing the Use of Depth Data for Fall Detection. *Personal Ubiquitous Comput.* 17, 6 (aug 2013), 1063–1072.

[28] Chaua C Queirolo, Luciano Silva, Olga RP Bellon, and Mauricio Pamplona Segundo. 2009. 3D face recognition using simulated annealing and the surface interpenetration measure. *IEEE transactions on pattern analysis and machine intelligence* 32, 2 (2009), 206–219.

[29] Daniel A Reid, Sina Samangooei, Cunjian Chen, Mark S Nixon, and Arun Ross. 2013. Soft biometrics for surveillance: an overview. *Handbook of statistics* 31 (2013), 327–352.

[30] Caroline Rougier, Edouard Auvinet, Jacqueline Rousseau, Max Mignotte, and Jean Meunier. 2011. Fall detection from depth map video sequences. In *International conference on smart homes and health telematics*. Springer, 121–128.