Thanh Nam Bach, Denise Junger*, Cristóbal Curio, Oliver Burgert

Towards Human Action Recognition during Surgeries using De-identified Video Data

De-identification Prototype for Visual Sensitive Information in the OR

https://doi.org/10.1515/cdbme-2022-0028

Abstract: With the progress of technology in modern hospitals, an intelligent perioperative situation recognition will gain more relevance due to its potential to substantially improve surgical workflows by providing situation knowledge in real-time. Such knowledge can be extracted from image data by machine learning techniques but poses a privacy threat to the staff's and patients' personal data. De-identification is a possible solution for removing visual sensitive information. In this work, we developed a YOLO v3 based prototype to detect sensitive areas in the image in real-time. These are then de-identified using common image obfuscation techniques. Our approach shows that it is principle suitable for de-identifying sensitive data in OR images and contributes to a privacy-respectful way of processing in the context of situation recognition in the OR.

Keywords: Action recognition, de-identification, sensitive information, image data, YOLO.

1 Introduction

Situation recognition in the operating room (OR) is one of the emerging fields in research. Most recent approaches use video data to recognize the actual surgical situation, e.g. the phase, during surgery [1]. Despite laparoscopic or microscopic videos, also an OR camera can be used as input for situation recognition. Such cameras can be attached e.g. to the ceiling, covering parts or the whole surgical area, to detect persons and their actions. In this case, sensitive data (e.g. visual personal identifier [2] or unstructured text data [3]) will be visible in the recorded video signal that enables the identification of persons. Processing such personal data, especially in the health section, need adequate protection. Via de-identification, all sensitive data can be made unrecognizable by reducing the association between identifying data and data subject [4]. Traditionally, obfuscation techniques are applied to irreversibly remove identifying data from the image, e.g removing the face. However, this has a negative impact on the usefulness of the image and may render the image useless for a situation recognition system.

To reduce the privacy risk of OR cameras without making the image useless for further processing, other de-identification techniques can be used. Several works already addressed de-identification methods, e.g general de-identification for visual personal identifiers [2] or faces [5]. Visual personal identifiers can be biometric (e.g. face, ear, gait, iris), soft-biometric (e.g. gender, age, tattoos), or non-biometric (e.g. text, license plate, hairstyle) [2].

In this work, a de-identification prototype is developed, which reduces the privacy risk by anonymizing a set of visual personal identifiers in video data. For that, a de-identification pipeline was established, including a camera system setup, a dataset for training and evaluation, as well as an anonymization algorithm.

2 Methods

2.1 Requirements analysis and risk assessment

A requirements catalog, identifiability assessment, and risk assessment were done for analysis. For requirements analysis, the vision and goals were defined for the de-identification prototype. The functional and non-functional requirements (FR/NFR) cover the following areas:

- FR1: Identify sensitive content
- FR2: De-identify sensitive content
- NFR1: Irreversibility
- NFR2: Intelligibility
- NFR3: Real-time processing
- NFR4: Accuracy

***Corresponding author: Denise Junger:** Reutlingen University, School of Informatics, Research Group Computer Assisted Medicine (CaMed), Reutlingen, Germany, e-mail: denise.junger@reutlingen-university.de
Thanh Nam Bach, Cristóbal Curio, Oliver Burgert: Reutlingen University, School of Informatics, Research Group Computer Assisted Medicine (CaMed), Reutlingen, Germany

To categorize the visual identifiers in the OR, an identifiability assessment was conducted. For each visual identifier, it was determined how well a human or algorithm unambiguously identifies a person based on that identifier, which resulted in an identifiability score (adapted based on [3]). The human component is based on subjective and unstructured observations of orthopedic surgeries, whereas the algorithmic component is based on broad literature research (e.g. [2]) and (if existing) identification solutions. The results are shown in Fig. 1.

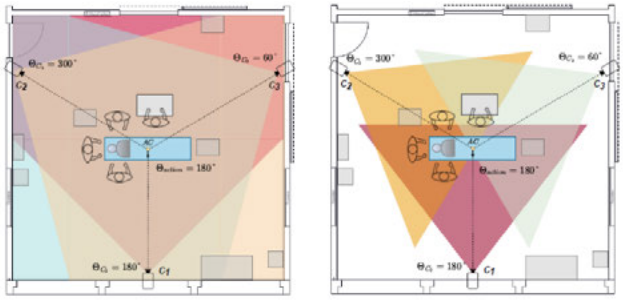
Identifier	Category	Identifiability			Modality			
		Hu	*Alg*	Score	RGB	Depth	IR	Skeletal data
Face	B	x	E	4	x	x	~	-
Gait	B	~	A	3	x	x	x	x
Ear	B	~	A	3	x	x	~	-
Body silhouette	SB	~	R/N	1-2	x	~	x	-
Gender, Age, Race, Ethnicity	SB	~	A	2	x	-	-	-
Scars, Marks, Tattoos	SB	x	A	2-4	x	-	~	-
Text	NB	x	E	0-4	x	-	~	-
Hairstyle, Dressing style	NB	~	R/N	1	x	-	~	-
Volumetric head image	B	-	-	unclear	x	-	~	-

Figure 1: Results of the identifiability assessment. NSB = Non-, Soft- or Biometric; Hu = Human: x = Commonly used, ~ = Not reliable, - = Not used; Alg = Algorithm: E = Established, A = Advanced, R/N = Research/Non existant; Score: 0 = Unrelated data, 4 = Data linked to individual; Modality: : x = Available, ~ = Available in lower quality, - = Not available.

For risk assessment, the data protection impact assessment (DPIA) was used to get a clear idea of the privacy risk for the OR scenario as well as be aware of the effectiveness and limits of de-identification of raw OR images. Therefore, the systematic description of the processing and its purposes, the assessment of the necessity and the proportionality of processing, and the risks concerning the rights and freedoms of the data subjects were conducted. The risk assessment states that de-identification measurements on visual data reduce the likelihood of the re-identification or prevent easy interpretation of the identifier, however, they do not prevent re-identification. Based on the exchange with data protection experts, the main concern lies with a re-identification attack that uses context information (like an OR plan), which can be successful even if the image itself is de-identified.

2.2 System architecture and setup

For the de-identification prototype, the camera system setup (see Fig. 2) for reliably acquiring actions during surgery was defined. The Azure Kinect RGB-D camera was used for this work. With the camera system, realistic visual data of the OR, containing human (inter)actions and sensitive information, was acquired as a basis for the de-identification prototype and annotated for training and testing purposes using bounding boxes of the category face, person, monitor, and badge. The dataset was acquired from three different angles using a single static camera repositioned at three different positions. For later usage, three cameras should be used in parallel. Five subjects performed multiple actions in the research OR at the same time and under different lighting conditions, wearing surgical clothes, facemasks, and hair nets (see Fig. 3). Due to the small resolution of the badges (< 16 mm per pixel) that was noticed within the first analyzed frames the badge annotations were removed from the dataset and not considered anymore for detection.



(a) Colour coverage (b) Depth coverage

Figure 2: Camera concept showing the three camera positions covering colour and depth.

The de-identification concept was defined and the anonymization pipeline implemented. The system architecture is depicted in Fig. 4. The prototype has two interfaces: the image sensors as input and the situation recognition system, which receives the output of de-identified images. The prototype consists of four components: The controller which manages the logic of the application, the image handler which buffers and provides the application with the image data, the detector which detects the sensitive regions of interest (ROI) in an image, and lastly the de-identifier which removes identifying information in an ROI.

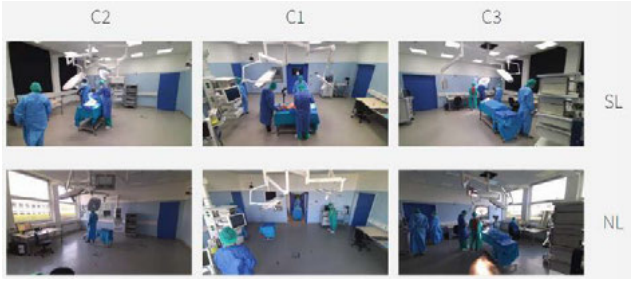


Figure 3: Dataset acquisition. C1-3 = Camera positions; SL = Synthetic lighting, NL = Natural lighting.

To identify sensitive content (i.e. the categories face, person, and monitor), a deep learning approach based on YOLO (You Only Look Once) v3 by [6] was implemented, detecting sensitive ROIs via object recognition. YOLO is an end-to-end network that inputs the whole image through a single pass and outputs the bounding boxes with class

probabilities, using a fixed set of candidate regions or so-called anchor boxes. As a basis, an out-of-the-box pre-trained network on the OpenImages dataset [7] (600 classes) was used and fine-tuned via the acquired dataset. For de-identification (see Fig. 5), common redaction methods of Masking, Blurring, and Pixelating were applied to the identified ROIs. These methods are the "naive" de-identification methods that only hinder human recognition and do negatively impact the intelligibility of the image.

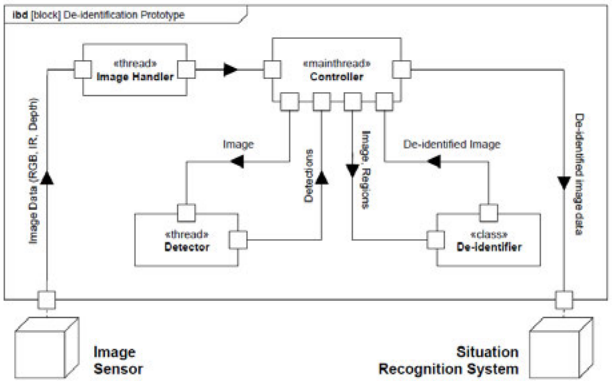


Figure 4: Block diagram of the de-identification prototype.

For evaluation, different YOLO models and training approaches were applied. The acquired dataset was used to assess the precision and performance of the anonymization. Furthermore, external image sources that differ in e.g. lighting and camera angle were tested to assess the transferability of the trained model to other circumstances.

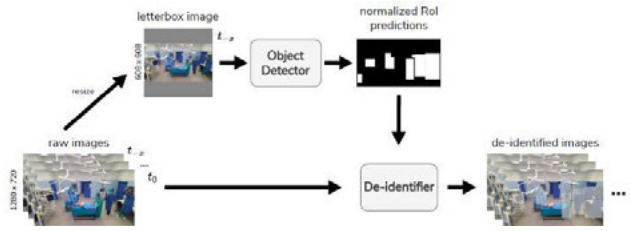


Figure 5: Simplified de-identification prototype concept.

3 Results

3.1 Training setup and dataset evaluation

The training using a fine-tuning approach with the standard YOLO v3 model and a learning rate of 0.001 created the best model with a recall of 90.7% and an mAP of 85.7% on the validation set. Based on the training results and the evaluation of the dataset, the validity of the camera setup is shown. Furthermore, the camera can sufficiently extract skeletal data from an OR scene, but with the limitation of two cameras at the same time because of the camera setup.

3.2 De-identification

By using parallelism, the de-identification prototype can achieve a required processing speed of 30 FPS. Given a detector speed of 13-14 FPS, every 3rd frame the detection is refreshed, which is sufficiently fast for the prototype. The de-identification takes place on all frames, based on the bounding box calculated every third frame. The prototype applies the basic de-identification methods to the image, which removes the sensitive information of the classes face, person, and monitor, but does negatively impact the intelligibility of the image. An example of the implemented face de-identification is shown in Fig. 6.

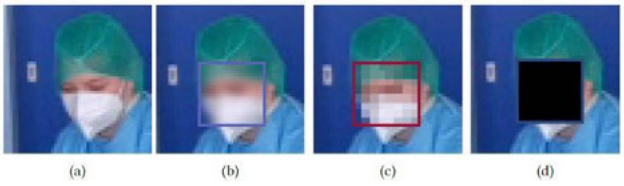


Figure 6: De-identification examples. (a) original, (b) blurring, (c) pixelating, (d) masking.

The de-identification prototype can successfully detect sensitive ROIs in our dataset. The evaluation with extern image sources not included in the training dataset showed that the model is also useful when using it with similar OR imagery which indicates the generalizability of the model. Nevertheless, the de-identification prototype acts differently on external images. On the one hand, the algorithm did very well in successfully detecting unmasked faces and persons in some examples. In some samples, the algorithm performs poorly on recognizing persons in the image, recognizing faces twice, with an inner and outer bounding box. Two examples showed unsuccessful and rather inaccurate detection results: Only half of the present faces were detected by the model with sufficient confidence. Furthermore, the algorithm is quite inaccurate in detecting monitors, not detecting nearly all available monitors in the image. Misclassifications also occurred, e.g. parts of a black chair were wrongfully recognized as monitors.

4 Discussion

This work addresses the need to automatically remove identifying information from image data of the OR. It shows

how sensitive data can be de-identified during surgery in the context of a situation recognition system.

The results from the dataset indicate that the camera setup adequately records the OR areas from multiple angles and the data can be used to implement a detector for sensitive objects. Using the YOLO v3 model and pre-trained weights an accurate and fast object detector for detecting sensitive objects in the OR was implemented. The trained model was also applied to several data instances, not present in the validation dataset. We assume that the performance is dependent on the difference in the object's sizes/appearance from the representations in the learned dataset. The reliability of the detector varies on external, unseen image data, especially when OR images are shot from a significantly different angle than the trained dataset. The poorer performance in detecting monitors is also very likely because of the difference in the trained dataset. Nevertheless, good generalization capabilities of the model and the validity of this approach for OR imagery could be shown overall. To gain better generalization capabilities, the dataset requires different class appearances and, therefore, needs to be extended to cover more conditions.

The implemented basic de-identification measurement achieved sufficient obfuscation in the image, that fulfills the anonymity attribute (without context information) based on DIN EN 62676-4. The aforementioned components are combined into the implementation of the de-identification prototype, which can de-identify a 30 FPS OR image stream with good accuracy. The detection speed of 14 FPS is sufficient for our case, given the assumption that no abrupt movements occur in the normal course of events in the OR.

Overall, our work shows the feasibility of such de-identification methods, although the current prototype lacks sufficient accuracy for clinical use. Therefore, further improvements are needed, e.g. by dealing with dataset imbalances in terms of class instances or object instance diversities. Furthermore, additional subjects should be included in the dataset. Ideally, real OR setups should be used.

5 Conclusions

Overall, the results from this work demonstrate a comprehensive implementation of a de-identification prototype. It is an important step for more privacy-respecting processing and enables future research to better take advantage of surgical data and pipelines. The chosen approach of using a pre-trained model and fine-tuning shows the effectiveness of deep learning approaches to detecting (sensitive) objects in an image. YOLO v3 specifically is a valid model for detecting sensitive areas in OR imagery and the trained model can be improved for future research. Furthermore, the de-identified video signal can be used for situation recognition in the OR, e.g. within the system of [8], but further improvements are needed to achieve the quality required in a practical environment.

Author Statement

Research funding: This research was funded by the Ministry of Science, Research and Arts Baden-Württemberg and the European Fund for Regional Development (EFRE).
Conflict of interest: Authors state no conflict of interest.
Informed consent: This article does not contain patient data.
Ethical approval: This article does not contain any studies with human participants or animals performed by the authors.

References

[1] Junger D, Frommer SM, Burgert O. State-of-the-art of situation recognition systems for intraoperative procedures. Med Biol Eng Comput. 2022;60(4):921-939. doi: 10.1007/s11517-022-02520-4.

[2] Ribaric S, Ariyaeeinia A, Pavesic N. De-identification for privacy protection in multimedia content: A survey. Signal Processing: Image Communication. 2016;47:131-151. doi: 10.1016/j.image.2016.05.020.

[3] Garfinkel SL. De-identification of personal information. National Institute of Standards and Technology (NIST). 2015. doi: 10.6028/NIST.IR.8053.

[4] ISO. ISO 25237:2017(en) Health informatics Pseudonymization. https://www.iso.org/obp/ui/#iso:std:iso:25237:ed 1:v1:en, accessed: 13.05.2022.

[5] Ren Z, Jae Lee Y, Ryoo MS. Learning to anonymize faces for privacy preserving action detection. Proceedings of the European Conferenc e o n Computer Vision (ECCV). 2018:620-636.

[6] Redmon J, Farhadi A. YOLOv3: An Incremental Improvement. 2018. https://arxiv.org/pdf/1804.02767, accessed: 30.06.2022.

[7] Kuznetsova A, Rom H, Alldrin N. et al. The Open Images Dataset V4. Int J Comput Vis. 2020;128:1956–1981. doi: 10.1007/s11263-020-01316-z.

[8] Junger D, Hirt B, Burgert O. Concept and basic framework prototype for a flexible and intervention-independent situation recognition system in the OR. Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization. 2021;1-6. doi: 10.1080/21681163.2021.2004446.