I C E M I T
1st International Scientific Conference on Economy, Management and Information Technologies – ICEMIT 2023

Architecture of an IP camera system using machine learning for privacy protection

Nikola Nikolić[a], Oliver Popović[a*], Vladica Ubavić[b], Marina Jovanović-Milenković[c]

[a] Toplica Academy of Applied Studies, Department of Business Studies Blace, Serbia
[b] Republic Geodetic Authority, Belgrade
[c] Educons University, Faculty of Project and Innovation Management, Belgrade

Article info

Review paper

DOI:
https://doi.org/10.46793/ICEMIT23.317N

UDC/ UDK:
004.738.5:342.721
004.738.5:343.45

Abstract

The development of technologies and the Internet conditioned the development of video surveillance systems. In modern video surveillance systems, the complete content recorded by a camera can be publicly available, i.e., visible on the Internet. The mass use of this type of camera has led to the fact that the cameras record everything and everyone and that the recorded content violates privacy. In order to protect privacy, different types of technologies and software are applied. One of those technologies is machine learning. By applying machine learning, it is possible to mask people's faces on recorded material. In this way, in publicly available content, as well as in case of unauthorized access, the identity of people is not compromised. This paper describes the architecture of a system that enables the application of machine learning for the purpose of protecting the identity of a person, where only an authorized person would have access to the original unmasked content.
Keywords: *machine learning, IP camera, video surveilance, privacy protection*

1. Introduction

The original application of the video surveillance system was physically limited by the object it was monitoring, so that the recorded material could not be seen outside the room being recorded.

IP Camera is created in response to the problem of the demand for improving image quality, efficiency, commercialization and enabling the wider use of video surveillance systems.

In addition to the mentioned improvements, the main improvement relates to making the complete content recorded by the camera or video surveillance system publicly available, i.e., visible on the Internet. The content recorded by the cameras can be accessed from many different devices such as computers, tablets and mobile devices. The mass use of this type of camera has led to the fact that the cameras record everything and everyone and that the recorded content is publicly available (example: cameras for recording public space or unsecured cameras). Violating the privacy of private persons is one of the consequences of this application of video surveillance cameras.

There are currently many tools and applications available on the market with functionalities that allow access to such cameras, recording content, etc. In order to protect personal privacy, various types of technology and software are applied, such as a technology called machine learning.

Machine learning is a field devoted to understanding and building methods that allow machines to "learn" - methods that use data to improve a computer's performance on some set of tasks. However, the use of machine learning (ML) in surveilance procedures is not an entirely new approach.

*Corresponding author
E-mail address: opopovic@gmail.com

The proposed architecure of application in this paper enables the masking of people's faces on recorded material. In this way, in publicly available content, as well as in the case of unauthorized access, the identity of people is not compromised.

2. Machine learning and facial recognition

Machine learning plays a key role in the field of facial recognition and video processing. The application of machine learning in these areas opens up many possibilities and provides significant benefits (Jordan et al, 2015).

Facial recognition is the process of identifying and verifying people based on their physical facial characteristics. Machine learning is used to train algorithms and models that can recognize and classify faces based on learned patterns. This is very useful in security systems, surveillance, digital authentication and other areas where it is necessary to identify individuals based on their faces.

Video processing also plays a significant role in machine learning. Machine learning enables automatic analysis and interpretation of video content. For example, machine learning algorithms can detect objects, track their trajectories, classify activities, and predict future events based on learned patterns. This has applications in various fields such as video surveillance, autonomous vehicles, medicine and many others.

Machine learning in video requires a large dataset to train models and algorithms. This data includes examples of faces, videos, movement patterns. Machine models then use this data to train and learn patterns, allowing them to recognize and analyze similar patterns in new video materials.

The process of identifying faces on video material includes several technical steps and algorithms:

1. **Face detection**: The first step is facing detection in the video material. This step involves searching for regions in the video that represent faces. Object detection algorithms, such as Haar cascade classification (Viola & Jones, 2001) or convolutional neural networks (Aurelien, 2019), are used for this.
2. **Face Isolation and Flattening**: After face detection, the next step is facing isolation and face flattening. This step aims to normalize the faces within the video material and allows for a consistent analysis. During flattening, faces may be reoriented, oriented, or rotated so that all faces in the video are in a consistent orientation.
3. **Facial Feature Extraction**: After face isolation and flattening, the next step is facial feature extraction. This step involves converting the faces into numerical feature vectors, which represent the internal characteristics and features of the face. Various techniques are used to extract facial features, including Faces with LBP (Local Binary Patterns) (Rahim et al., 2013), Gabor filters (Fogel & Sagi 1989), and convolutional neural networks (Aurelien, 2019).
4. **Classification and identification**: The final step is the classification and identification of the face. Based on the facial feature vectors, a machine learning model or algorithm is used to classify and identify faces. Different methods are used for this step, including powerful classifier methods such as Support Vector Machines (SVM), Random Forests or gradient trees.

This process of identifying faces on video material can be performed in real time or in post-processing, depending on the requirements and needs of the application. It is important to keep in mind that there are many factors that can affect the success and performance of facial recognition, such as lighting, facial pose, appearance changes, etc. That is why continuous research and development in this area are important for achieving more accurate and efficient methods of identifying faces on video material.

3. Methods of implementing video obfuscation

Some of the papers in the field of visual obfuscation reviewed in this work are listed in Table 1. Importance is given to research published in the field of perceptual obfuscation, as it is especially relevant for this paper.

Table 1. Some types of image filtering

Image Filtering	Reference
Morphing	Yeh et al. (2017), Semantic image inpainting with deep generative models
Warping	Yu et al. (2018), Generative image inpainting with contextual attention
Cartooning - using convolutional neural networks	Rong et al. (2021), FrankMocap: A monocular 3D whole-body pose estimation system via regression and integration
Adversarial stickers and patches	Kapadia et al. (2007) Virtual walls: Protecting digital privacy in pervasive environments Hoory et al. (2020), Dynamic adversarial patch for evading object detection models

In morphing image filtering, Yeh et al. (2017) described semantic image inpainting, where missing regions have to be filled based on the available visual data. They described a novel method for semantic image inpainting, which generates

the missing content by conditioning on the available data. A trained generative model is given, where the search for the closest encoding of the corrupted image is conducted.

Warping image filtering (Yu et al, 2018) is conducted by deep generative model-based approach, which can synthesize novel image structures and explicitly utilize surrounding image features as references during network training to make better predictions. The model is a feed-forward, fully convolutional neural network which can process images with multiple holes at arbitrary locations and with variable sizes during the test time.

Cartooning is advanced image filtering approach. Rong et al (2021) described an implementation called *FrankMocap*. It represents a whole-body 3D pose estimation system that can produce 3D face, hands, and body simultaneously from monocular images. The core idea of FrankMocap is its modular design: three different integration modules that trade off between latency and accuracy. All of them are capable of providing solutions to unify the separate outputs into seamless whole-body pose estimation results.

Kapadia et al. (2007) addressed the problem of privacy problems in the digital world, and recomended implementation of policy language based on the metaphor of physical walls, and posit that users will find this abstraction to be an intuitive way to control access to their digital footprints.

Hoory et al. (2020) investigated the application of neural networks for object recognition, as well as possibilities for avoiding recognition (e.g., a carefully crafted sticker placed on a stop sign). Furthermore, they presented an innovative attack method against object detectors applied in a real-world setup, that addresses some of the limitations of existing attacks. This method uses dynamic adversarial patches which are placed at multiple predetermined locations on a target object.

In scenarios where the target of obfuscation is human observers, the objective of various techniques is to provide visual privacy to individuals who seek to shield themselves from prying eyes lacking the requisite access privileges. This particular category of methods primarily strives to generate images where elements bearing privacy sensitivity are rendered perceptually distinct from their original counterparts.

Perceptual obfuscation methods can exhibit different characteristics. They can either be reversible, wherein the original image can be (1) recovered following modification, (2) retrieved if separately recorded, or (3) rendered irreversible. A comprehensive overview of the classical literature on perceptual obfuscation is thoughtfully presented by Padilla-López et al (2015).

For the specific realm of facial anonymization, morphing and warping serve as indispensable filtering techniques. In morphing, the input facial features seamlessly transform into those of a designated target face. This transformation is orchestrated through interpolation and intensity adjustments, finely tuning the positions of facial keypoints to align with the target visage. In warping, an array of keypoint parameters is ascertained through adept face detection techniques. Subsequently, these keypoints are judiciously shifted in response to a 'warping strength' parameter, resulting in novel intensity values determined via interpolation.

The concept of cartooning has emerged within the literature as a viable approach for image filtering with privacy implications. Through this method, the total color palette is significantly reduced, and image textures are simplified based on the properties of neighboring pixels. Additionally, edge recovery techniques are employed to retain the sharpness of edges in the image.

In a different vein, the realm of adversarial stickers and patches methods hinges on the creation of adversarial examples through the strategic deployment of physical artifacts. When exposed to cameras during capture, these artifacts increase the likelihood of the subject being erroneously identified, effectively enhancing the odds of misidentification. This and many more approaches related to visual privacy preservation are thoroughly reviewed by Ravi & Climent-Pérez (2023).

Furthermore, the work of Zhou and Pun (2021) introduces a novel dimension with their creation of 'Face Pixelation in Video Live Streaming.' This innovative approach enables real-time face tracking and pixelation, a development of significant interest. The authors of this paper are actively pursuing the integration of this approach into their own system architecture, signaling exciting possibilities in the field.

4. Proposed architecture

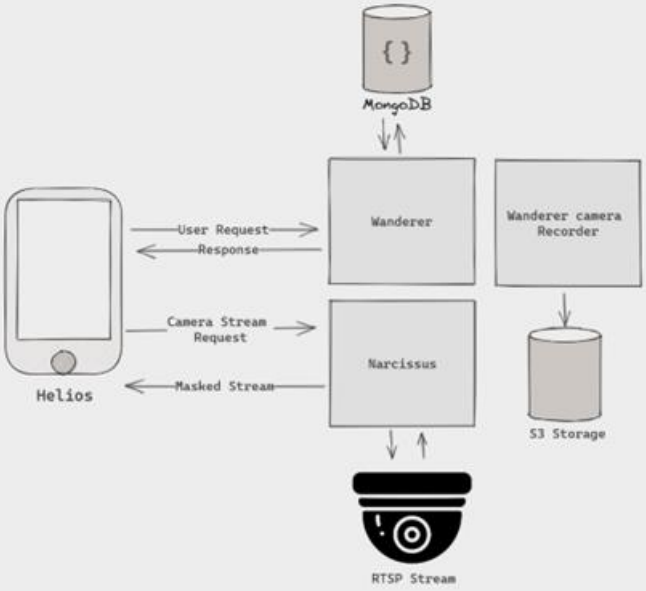
As part of the project, focus was on the technological aspects of face identification on video footage. This demanding process involves the application of advanced machine learning and computer vision algorithms.

Proposed architecture will provide a wide range of functionality that includes allowing a masking the faces of one or more persons on one or several cameras.

The project consists of 3 main parts, which are as follows:

1. "Narcissus" - Python gRPC server that uses machine learning library to transform video and mask people's faces.
2. "Wanderer" - Kotlin gRPC server that has the function of user authentication and resource management, as well as camera recording.
3. "Helios"– Jetpack Kotlin application that provides an interface to the user for controlling the camera(s).

Figure 1. Diagram of proposed architecture of an IP camera system (Source: Authors)



As part of the system, MongoDB database and Amazon S3 cloud system will be used for data storage.

MongoDB is a document-oriented database that is popular in the world of software development and data processing. This database is designed to be flexible, scalable and easy to use. MongoDB is a popular choice for many types of applications, including web applications, microservice architectures, mobile applications, and analytics systems. It enables fast utilization of data and enables flexibility in developing and upgrading applications.

Amazon S3 (Simple Storage Service) is a cloud storage service provided by Amazon Web Services (AWS). It enables companies and individuals to store, protect and access their data on the Internet through secure and scalable servers. Amazon S3 has become a popular choice for many organizations and applications worldwide due to its scalability, availability, and security. It provides a platform for efficient and secure cloud data storage and management.

5. Discussion and conclusion

The advancement of technology has played a pivotal role in shaping the evolution of video surveillance systems. In contemporary video surveillance systems, cameras have the capability to make their entire recorded accessible via the Internet. The widespread adoption of such cameras has resulted in the indiscriminate recording of various individuals and activities, often infringing upon personal privacy rights. To address these concerns and safeguard privacy, a diverse range of technological solutions, including machine learning, have been employed.

Machine learning, as a crucial technology in this context, offers the means to obfuscate the identities of individuals captured in recorded footage. This process entails concealing or altering facial features by morphing, warping, cartooning or bluring part of the videos, thereby ensuring that the identity of individuals remains protected in publicly accessible content, especially in instances of unauthorized access. This paper delineates the architectural framework of a system designed for the application of machine learning techniques, specifically geared towards safeguarding the anonymity of individuals. The system's design ensures that only authorized individuals possess the means to access the unaltered, unmasked content.

The basic idea is that the proposed system has two streams: one with a processed signal in real time, where the faces of the persons will be masked, and the other unmasked, which will be stored separately, where only authorized persons will have access.

For future work, the development of the proposed system is foreseen, with the implementation of all the described functionalities.

References

Aurélien, G. (2019). *Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow.* O'Reilly Media.

Fogel, I., & Sagi, D. (1989). Gabor filters as texture discriminator. *Biological cybernetics*, *61*(2), 103-113. https://doi.org/10.1007/BF00204594

Hoory, S., Shapira, T., Shabtai, A., & Elovici, Y. (2020). Dynamic adversarial patch for evading object detection models. *arXiv preprint arXiv:2010.13070*.

Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, *349*(6245), 255-260. https://doi.org/10.1126/science.aaa8415

Kapadia A, Henderson T, Fielding JJ, Kotz D (2007) Virtual walls: Protecting digital privacy in pervasive environments. In LaMarca A, Langheinrich M, Truong KN (eds) *Pervasive computing* (pp. 162–179) Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-72037-910

Ravi, S., Climent-Pérez, P., & Florez-Revuelta, F. (2023). A review on visual privacy preservation techniques for active and assisted living. *Multimedia Tools and Applications*, 1-41. https://doi.org/10.1007/s11042-023-15775-2

Padilla-López, J. R., Chaaraoui, A. A., & Flórez-Revuelta, F. (2015). Visual privacy protection methods: A survey. *Expert Systems with Applications*, *42*(9), 4177-4195. https://doi.org/10.1016/j.eswa.2015.01.041

Rahim, M. A., Hossain, M. N., Wahid, T., & Azam, M. S. (2013). Face recognition using local binary patterns (LBP). *Global Journal of Computer Science and Technology*, *13*(4), 1-8.

Rong, Y., Shiratori, T., & Joo, H. (2021). Frankmocap: A monocular 3d whole-body pose estimation system via regression and integration. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 1749-1759). https://doi.org/10.1109/iccvw54120.2021.00201

Viola, P., & Jones, M. (2001, December). Rapid object detection using a boosted cascade of simple features. In *Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR 2001* (Vol. 1, pp. I-I). Ieee. https://doi:10.1109/cvpr.2001.990517

Yeh, R. A., Chen, C., Yian Lim, T., Schwing, A. G., Hasegawa-Johnson, M., & Do, M. N. (2017). Semantic image inpainting with deep generative models. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 5485-5493). https://doi.org/10.1109/cvpr.2017.728

Yu, J., Lin, Z., Yang, J., Shen, X., Lu, X., & Huang, T. S. (2018). Generative image inpainting with contextual attention. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 5505-5514). https://doi.org/10.1109/cvpr.2018.00577

Zhou, J., & Pun, C. M. (2020). Personal privacy protection via irrelevant faces tracking and pixelation in video live streaming. *IEEE Transactions on Information Forensics and Security*, *16*, 1088-1103. https://doi.org/10.1109/TIFS.2020.3029913