DEVELOPMENT OF AN IP CAMERA SYSTEM USING MACHINE LEARNING FOR THE PURPOSE OF PRIVACY PROTECTION

Oliver Popović[1], Nikola Nikolić[1], Vladica Ubavić[2], Marina Jovanović-Milenković[3], Marko Asanović[4], Ivana Buzdovan[4]

[1] Toplica Academy of Applied Studies, Prokuplje
[2] Republic Geodetic Authority, Belgrade
[3] Faculty of Project and Innovation Management, Educons University, Belgrade
[4]Faculty of Transport, Communications and Logistics, Budva

Abstract

The rapid evolution of technology and the Internet have profoundly influenced the advancement of video surveillance systems. In contemporary video surveillance systems, camera-recorded content has the potential to be publicly accessible via the Internet. However, the widespread deployment of such cameras has raised concerns about privacy violations, as these devices capture extensive footage of individuals and activities. To address these concerns and safeguard privacy, various technologies and software solutions have been deployed. This paper elaborates the architectural design and development of a system that use machine learning techniques to obfuscate individuals' faces within recorded materials. By employing machine learning, the proposed system ensures that the identity of individuals remains protected, both in publicly available content and in cases of unauthorized access, granting exclusive access to the original, unobscured content only to authorized personnel.

Keywords: machine learning, IP camera, video surveilance, privacy protection.

INTRODUCTION

The initial purpose of video surveillance systems was constrained by the physical limitations of the monitored objects, restricting access to the recorded footage beyond the monitored area.

The development of IP Cameras emerged as a solution to address various challenges, including the need for enhanced image quality, operational efficiency, commercial viability, and the broader adoption of video surveillance systems. An essential aspect of this advancement was the capability to make the recorded content accessible to the public via the Internet. Users could now access camera-recorded content from a variety of devices, such as computers, tablets, and mobile devices. Widespread usage of these cameras, especially in public spaces and unsecured locations, resulted in the unintentional invasion of privacy for individuals (Kalbo, N. et al. 2020).

As a response to these challenges, a multitude of tools and applications have been developed in the market and scientific community, offering functionalities to access, record, and manage such cameras (Elharrouss, O. et al. 2021). Most solutions emphasize automatic face recognition, object and violence detection as their advantage (Ye, M. et al. 2021. Elhoseny, M. 2020. Sreenu, G., & Durai, S. 2019. Omarov B. et al. 2022.).

Machine learning and data learning, a fields dedicated to developing methods that allow machines to learn and improve their performance through data analysis, has also found applications in surveillance (Tsakanikas, V., & Dagiuklas, T. 2018. Xu, J. 2021).

Regrettably, authors noted that there is a notable absence of functionalities designed to safeguard privacy in relation to recorded material.

The application examined in this paper, which leverages machine learning techniques, facilitates the obfuscation of individuals' facial features in recorded material. Consequently, both in publicly accessible content and instances of unauthorized access, the integrity of individuals' identities remains safeguarded.

FACE DETECTION AND MACHINE LEARNING

Machine learning assumes a pivotal role in the domains of facial recognition and video processing, offering substantial advantages and diverse applications (Jordan et al., 2015).

Facial recognition involves the identification and verification of individuals based on their distinct facial features. Machine learning is employed to train algorithms and models that can discern and categorize faces by recognizing underlying patterns. This technology finds application in security systems, surveillance, digital authentication, and various scenarios where facial identification is essential.

In video processing, machine learning plays an equally significant role. It empowers the automated analysis and interpretation of video content. For instance, machine learning algorithms are adept at identifying objects, tracking their movements, classifying activities, and even predicting future events based on learned patterns. These capabilities have far-reaching applications across domains such as video surveillance, autonomous vehicles, healthcare, and more.

However, successful implementation of machine learning in video processing necessitates extensive datasets for training models and algorithms. These datasets encompass samples of faces, videos, and movement patterns, enabling machine models to learn and recognize similar patterns in new video materials.

The process of facial recognition in video materials involves several technical steps and algorithms:

Face Detection: Initial face detection is crucial in video materials. This phase entails 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 commonly used for this purpose.

Face Isolation and Flattening: Following face detection, the next step is face isolation and flattening (Chihaoui M. et al. 2016). This step aims to standardize the orientation and presentation of faces within the video material, ensuring consistent analysis. Faces may be reoriented, aligned, or rotated to maintain uniformity.

Facial Feature Extraction: Subsequent to face isolation and flattening, facial feature extraction comes into play. This process involves translating facial data into numerical feature vectors, encapsulating the internal characteristics and attributes of each face. Diverse techniques are employed for feature extraction, including Local Binary Patterns (LBP) for faces (Rahim et al., 2013), Gabor filters (Fogel & Sagi, 1989), and convolutional neural networks (Aurelien, 2019).

Classification and Identification: The final phase revolves around classifying and identifying faces based on the extracted facial feature vectors. Machine learning models or algorithms are leveraged for this purpose, with various methods available, such as Support Vector Machines (SVM), Random Forests, or gradient trees.

The process of identifying faces in video materials can be performed in real-time or during post-processing, depending on specific application requirements. It is essential to recognize that various factors, including lighting conditions, facial poses, and appearance variations, can influence the accuracy and efficacy of facial recognition. Consequently, ongoing research and development in this field are imperative to refine and enhance methods for identifying

faces in video materials, ensuring their precision and efficiency.

IP CAMERA SYSTEM ARCHITECTURE

The primary focus of this system is on the technological aspects of facial recognition in video footage, a complex process requiring the application of cutting-edge machine learning and computer vision algorithms.

The developed architecture (*Figure 1*) have capability to mask the faces of one or more individuals across one or multiple cameras (Nikolić et al, 2023).

The project comprises three key components:

"Narcissus" - This Python gRPC server harnesses a machine learning library to process video content and apply facial masking to safeguard the privacy of individuals.

"Wanderer" - Implemented in Kotlin, this gRPC server serves a multifaceted role. It handles user authentication and resource management while also managing camera recording.

"Helios" - An mobile application developed in Jetpack Kotlin, "Helios" provides an intuitive user interface for camera control, offering users the means to oversee one or more cameras seamlessly.

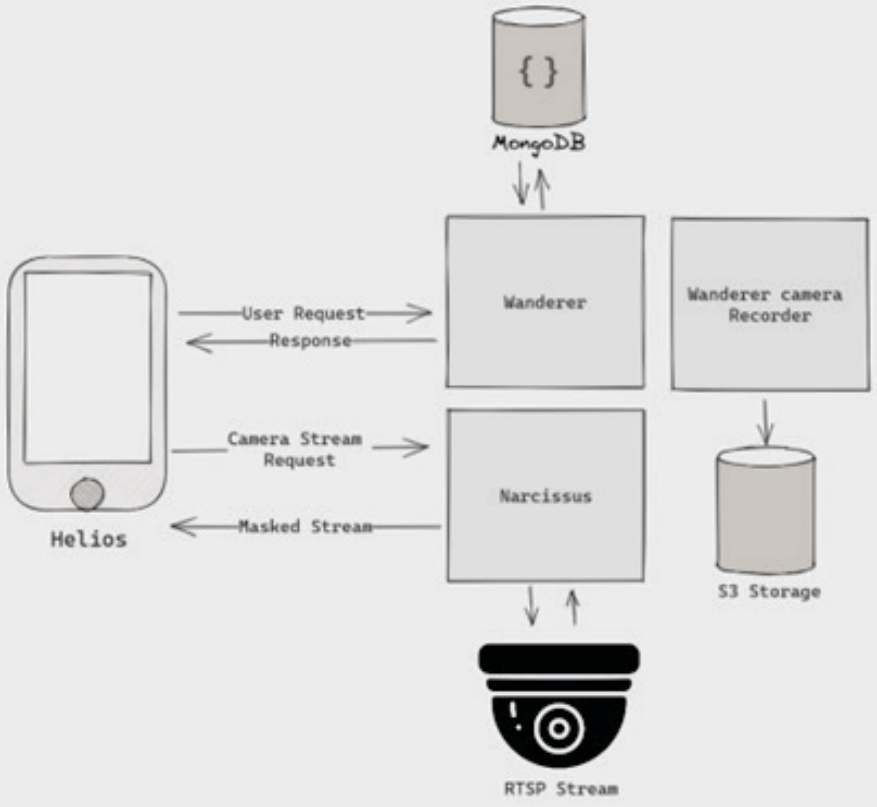

Fig. 1. *Diagram of architecture of an IP camera system*

To ensure secure communication, a JWT token is generated upon user sign-in and subsequently safeguarded on trusted devices to serve as a session credential. This token is indispensable for facilitating communication between Narcissus and Wanderer.

These services interoperate through various means. Helios employs generated gRPC client code to retrieve and store data, facilitating the exchange of user JWT tokens in the process.

Regarding the masked video, "Narcissus" encodes video into discrete units of bytes and transmits them using the GRPC Server streaming method. The stream is terminated when it is no longer required.

Together, these components form an integrated system designed to address the

intricate demands of face recognition and privacy protection in video content.

IP CAMERA SYSTEM DEPLOYMENT

The Wanderer application database consists of two collections:

- Users - Used to store user information.

- Cameras - Used to store all available cameras.

When the Wanderer application activates the camera recording, the video recording process begins by applying the H264 (MP4) video compression codec. This codec effectively compresses video content and ensures an optimal balance between video quality and storage space. Each of camera recording lasts 1 hour, where each day contains 24 recordings. This process results in a clean video with full quality.

Helios mobile application (*Figure 2*) respresents an user interface toward IP Camera system.

The application's core functionalities can be categorized as follows:

Camera Data Functionality: This encompasses tasks such as listing existing cameras, creating new ones, deleting cameras, and modifying existing entries. These actions are facilitated by the "Wanderer" gRPC client, which sends the user JWT token to establish a connection with the user and ensure proper access rights.

Camera Video Feed Functionality: This pertains to presenting the user with the camera feed. This process is orchestrated by "Narcissus," which transmits specific details via gRPC server streaming. Additionally, "Narcissus" returns secure videos (masked) to the user. Notably, the feed is restricted and does not permit advanced operations such as zooming or viewing specific timestamps.

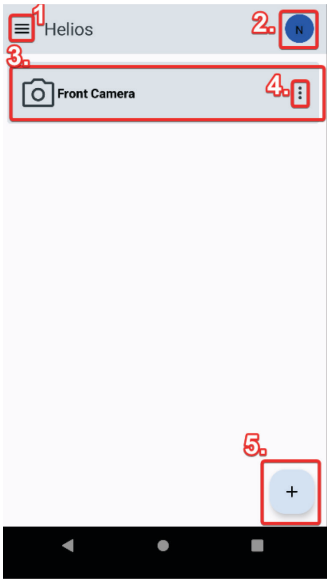

Fig. 2. *Home screen of Helios mobile application*

Content description:
1. Burger menu,
2. User profile,
3. Display of available cameras,
4. Options/Camera Settings,
5. Adding a new camera.

Selecting the camera tab configuration option initiates the display of a drop-down menu, affording the user the opportunity to modify the camera name, adjust the link for accessing the camera stream, and configure recording settings.

It is worth noting that the option to incorporate multiple cameras is available.

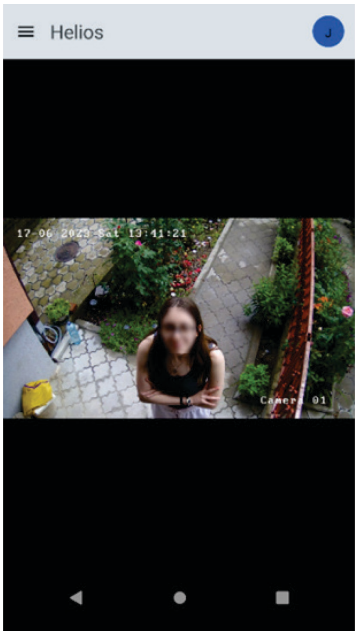

Fig. 3. Live *camera display with masked face*

When a person's face is detected on the camera, it is automatically masked by Narcissus (Figure 3). System has possibility to recognize and mask multiple faces simultaneously.

CONCLUSION

The considerable potential that applications offer acknowledges how the escalating ubiquity of cameras worldwide emphasizes the threat to privacy. It is imperative to bolster public awareness regarding the issue and significance of online privacy. This have relevance primarily to the safeguarding of one's identity and personal information. Ensuring the protection of identity should assume highest importance, particularly for applications involved in video surveillance.

Moreover, the challenge of object masking that this system addresses can be further extended to enhance the detection of a broader array of objects within images. This would serve to fortify the protection of not only facial features, but also other sensitive information such as vehicle registration markings and personal documents, among others.

Additionally, the IP camera system can be enhanced by incorporating additional functionality into the camera feed, thereby enabling more sophisticated operations such as zooming and the ability to view specific timestamps.

Enhanced security measures can be achieved by incorporating a refresh token in tandem with an access token, thereby mitigating potential vulnerabilities to JWT-based attacks.

A comparison of this developed IP camera system with others similar systems will be part of future work, because it requires a significant and wider survey.

REFERENCE

[1] Kalbo, N. Mirsky, Y. Shabtai, A. Elovici, Y. The Security of IP-Based Video Surveillance Systems. *Sensors*, *20*, 2020, p.4806.

[2] Elharrouss, O., Almaadeed, N., & Al-Maadeed, S. A review of video surveillance systems. Journal of Visual Communication and Image Representation, 77, 2021, 103116.

[3] Ye, M., Shen, J., Lin, G., Xiang, T., Shao, L., & Hoi, S. C. Deep learning for person re-identification: A survey and outlook. IEEE transactions on pattern analysis and machine intelligence, 44(6), 2021, p.2872-2893.

[4] Elhoseny, M.Multi-object detection and tracking (MODT) machine learning model for real-time video surveillance systems. Circuits, Systems, and Signal Processing, 39, 2020, p.611-630.

[5] Sreenu, G., & Durai, S. Intelligent video surveillance: a review through deep learning techniques for crowd analysis. Journal of Big Data, 6(1), 2019, p.1-27.

[6] Omarov B, Narynov S, Zhumanov Z, Gumar A, Khassanova M. 2022. State-of-the-art violence detection techniques in video surveillance security systems: a systematic review. PeerJ Computer Science 8, 2022, e920.

[7] Tsakanikas, V., & Dagiuklas, T. Video surveillance systems-current status and future trends. Computers & Electrical Engineering, 70, 2018, p.736-753.

[8] Xu, J. A deep learning approach to building an intelligent video surveillance system. Multimedia Tools and Applications, 80(4), 2021, 5495-5515.

[9] Jordan M. I., Mitchell T. M. Machine learning: Trends, perspectives, and prospects. Science, 349(6245), 2015, p.255-260.

[10] Viola P. Jones M. Rapid object detection using a boosted cascade of simple features. Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001. Vol. 1. IEEE Comput. 2001

[11] Aurélien G. Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow. Sebastopol, CA: O'Reilly Media. 2019.

[12] Chihaoui M, Elkefi A, Bellil W, Ben Amar C. A Survey of 2D Face Recognition Techniques. Computers. 2016; 5(4):21.

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

[14] Fogel, I., & Sagi, D. (1989). Gabor filters as texture discriminator. Biological cybernetics, 61(2), 1989, p.103-113.

[15] Nikolić N., Popović O., Ubavić V., Jovanović Milenković M. Architecture of an IP camera system using machine learning for privacy protection. 1st International Scientific Conference on Economy, Management and Information Technologies-ICEMIT, 2023.