A Virtualized Video Surveillance System for Public Transportation*

Talmaj Marinč[1], Serhan Gül[1], Cornelius Hellge[1], Peter Schüßler[2], Thomas Riegel[3], and Peter Amon[3]

[1] Fraunhofer Heinrich Hertz Institute, Einsteinufer 37, 10587 Berlin, Germany
{talmaj.marinc, serhan.guel, cornelius.hellge}@hhi.fraunhofer.de
[2] DResearch Fahrzeugelektronik GmbH, Otto-Schmirgal-Strae 3, 10319 Berlin, Germany schuessler@dresearch-fe.de
[3] Siemens Corporate Technology, Otto-Hahn-Ring 6, 81739 Münich, Germany
{thomas.riegel, p.amon}@siemens.com

Abstract. Modern surveillance systems have recently started to employ computer vision algorithms for advanced analysis of the captured video content. Public transportation is one of the domains that may highly benefit from the advances in video analysis. This paper presents a video-based surveillance system that uses a deep neural network based face verification algorithm to accurately and robustly re-identify a subject person. Our implementation is highly scalable due to its container-based architecture and is easily deployable on a cloud platform to support larger processing loads. During the demo, the users will be able to interactively select a target person from pre-recorded surveillance videos and inspect the results on our web-based visualization platform.

Keywords: video-based security, surveillance, face verification

1 Introduction

Intelligent surveillance systems are increasingly playing an important role in the identification of potential security threats in private and public spaces. The new generation of multimedia surveillance systems collects, stores, and analyzes information from various sensors, and implement advanced mechanisms for event notification and sharing. Computer vision based systems are rapidly gaining importance due to various reasons including the increased quality of the capture devices, increased processing capabilities enabled by the developments in graphics processing technologies, and the availability of public and private clouds providing massive amounts of computation power [4].

Several intelligent surveillance systems have been proposed in recent years, and some real-world deployments have also been reported [6]. Camps et al. [1] deployed a person re-identification system at a busy airport in the USA. Their

* This research has received funding from the German Federal Ministry for Economic Affairs and Energy under the VIRTUOSE-DE project.

system integrates various computer vision algorithms such as foreground detection, pedestrian detection as well as person tracking, and operates using the airport's network infrastructure in real time. Zhang et al. [7] present a real-time distributed wireless surveillance system for surveillance in enterprise campuses. Their system intelligently partitions the computing among the local device, different edge computing nodes, and the cloud. In this paper, we introduce a video-

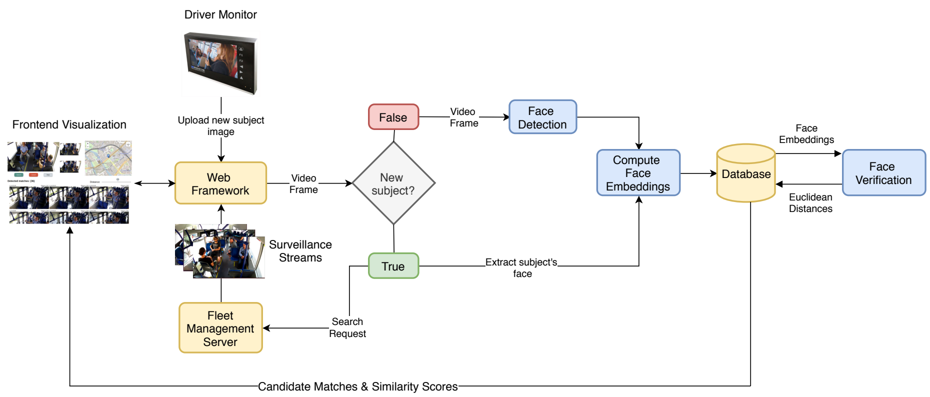

Fig. 1. System architecture and workflow of the proposed video surveillance system for public transportation.

based security system for public transportation (PT), specifically for surveillance in public buses. Our system utilizes the surveillance cameras installed in a bus and enables the re-identification of a suspect when (s)he switches to another vehicle within a pre-defined region of interest (RoI). The proposed system can be used by various stakeholders such as fleet management system (FMS) providers, technology providers (infrastructure as a service, IaaS) as well as governments and public authorities. Our system is *modular* in the sense that it contains easily exchangeable software components that communicate over a simple REST API. This enables easy upgrade of the video analysis components (e.g., neural network) when better performing algorithms become available. Secondly, our system is highly *scalable* due to its architecture that uses Docker containers for component-level virtualization. Thus, it is possible to easily start multiple instances of a processing block as well as efficiently manage several search processes through an orchestration software located in a cloud server. Finally, our system is capable of providing more advanced analysis compared to motion detection and background subtraction based systems, due to its advanced face verification module based on a pre-trained DNN.

2 System Overview

The proposed system comprises the following components which are displayed in Figure 1. Each component runs in its own Docker container enabling easy horizontal scalability of the processing blocks.

Web Framework. The backend logic is developed in Django, a web framework written in Python. For serving the static files we employ a more efficient Nginx server with a reverse proxy to the Django backend. For spawning conccurent tasks, we use the task manager Celery. A spawned task is terminated after a pre-defined timeout or may be killed directly through a post request.

Video Source. The public buses used in our setup are equipped with multiple IP cameras delivering resolutions up to 1080p and certified for usage in vehicles (on-board). The video streams from the cameras (H.264/AVC) are fetched by the Mobile Digital Video Recorder (MDVR, HydraIP MR4410) which provides the integrated 3G/4G/WLAN communication modules for mobile ground communication from vehicle to the cloud and/or back office application.

Fleet Management System (FMS). FMS is a web-based, multi-user back office application for centralized alarm management, diagnostic, and maintenance purposes of the on-board systems. The application allows controlling video streams from the video recorders of an entire bus fleet and includes features such as location management, map view, and management of vehicle metadata.

Face Detection. In order to detect the faces that serve as comparison images in the face verification process, we use the Single Shot Detector (SSD) [3] implemented in OpenCV. SSD relies on a pre-trained DNN for detection and the OpenCV implementation employs a ResNet [2] base network which provides high detection accuracy with very low false positive rates.

Face Verification. Faces are compared through the embeddings computed by a pre-trained Facenet [5] model from Dlib. It projects each face image onto a 128-dimensional vector space, in which the faces belonging to the same person are close to each other in terms of their Euclidean distances. Our system allows visual confirmation of the face verification output by a human operator in order to eliminate potential false positives before alarming the security personnel.

Message Queues and Database. We use the in-memory database Redis for queuing of the images that are to be processed by the face detection and face verification components. These processing components asynchronously pull and process the data. Since the algorithms are constantly loaded in memory, they are ready to immediately process any data that appear in the queues.

In order to store the relevant data, we use the database PostgreSQL. PostgreSQL allows saving the face embeddings, which are used for face verification, as arrays. Thus, the expensive operation of computing face embeddings is performed only once for each face image. Since the computation of Euclidean distance between embeddings is cheap, this setup enables fast comparison between any two subjects stored in the database.

Visualization. Although intelligent computer-based surveillance systems have reached amazing capabilities, humans are still necessary for approving further actions, especially if those concern fundamental rights. To support such a decision, a browser-based dashboard has been implemented, which allows a human operator in the control center to approve incoming notifications, trigger a search request, inspect retrieved results, accept one of the proposed matches, and terminate the search, or alternatively, reject the matches and continue the search.

3 Proof-of-Concept Setup

In this demonstration, we show how the system components work together to re-identify a searched subject in a surveillance network. For this, we use pre-recorded video streams from multiple cameras in two buses containing 11 different subjects. During the recording session, the subjects went in and out of the field-of-views of different cameras within a bus, and also changed from one bus to another. We use the video streams from Bus 1 to simulate a touchscreen where the user (in the role of the bus driver) can interactively choose the suspect to be searched. The video streams from Bus 2 simulate the surveillance streams uploaded by FMS that are used to re-identify the subject.

Our demo setup consists of two screens: one shows the video streams on which the simulation if performed, and the other shows the results on a web browser on our visualization dashboard. We show the initial subject selection, the detected faces on the surveillance videos and the candidate matches provided by the face verification model. Users may interact and select the search subject in the provided set of surveillance videos. Demo video is available at: `https://datacloud.hhi.fraunhofer.de/nextcloud/s/dX8ZLi7PRQ22YTA`.

References

1. Camps, O., Gou, M., Hebble, T., Karanam, S., Lehmann, O., Li, Y., Radke, R.J., Wu, Z., Xiong, F.: From the lab to the real world: Re-identification in an airport camera network. IEEE transactions on circuits and systems for video technology **27**(3), 540–553 (2017)
2. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778 (2016)
3. Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.Y., Berg, A.C.: Ssd: Single shot multibox detector. In: European conference on computer vision. pp. 21–37. Springer (2016)
4. Neal, D., Rahman, S.: Video surveillance in the cloud? arXiv preprint arXiv:1512.00070 (2015)
5. 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)
6. Ye, Y., Ci, S., Katsaggelos, A.K., Liu, Y., Qian, Y.: Wireless video surveillance: A survey. IEEE Access **1**, 646–660 (2013)
7. Zhang, T., Chowdhery, A., Bahl, P.V., Jamieson, K., Banerjee, S.: The design and implementation of a wireless video surveillance system. In: Proceedings of the 21st Annual International Conference on Mobile Computing and Networking. pp. 426–438. ACM (2015)