Evaluation of an Embedded Computer Vision System for Outdoor Media Audience Analysis

Bogusław Cyganek[1,2], Piotr Janyst[1], Mateusz Knapik[2], Łukasz Przebinda[1], Marta Ptaszkiewicz[1], and Tomasz Tajmajer[1]

[1] MyLED, Cracow, Poland {p.janyst, l.przebinda, m.ptaszkiewicz, t.tajmajer}@myled.pl
[2] AGH - University of Science and Technology, Cracow, Poland {cyganek, mknapik}@agh.edu.pl

Abstract. As digital advertisement screens become more popular, there is a need of providing technologies for monitoring the audience of such media. We present the ARA system, which consists of a hardware platform for GPU-accelerated computer vision and a SaaS platform for data analysis. We conducted experiments to evaluate the accuracy of the system in pedestrian counting scenario as compared to a human observer. In this paper we present architecture of this system, as well as the results of its evaluation.

Keywords: Pedestrian monitoring· Machine Learning · Audience analysis, Digital out-of-home media

1 Introduction

Recent growth of digital-out-of-home (DOOH) market result in higher demand for physical audience analysis technologies. The stakeholders are used to the customer data and ease of targeted marketing available in the main digital medium - the world wide web. Similar expectations arose for new kinds of media that are no longer purely digital - such as outdoor screens, interactive displays or augmented reality. Here we present a technology which aims to provide physical audience monitoring - dedicated for use with the DOOH media. The proposed ARA system is a mix of an AI-enabled computer vision hardware with an analytic SaaS platform. The basic functionality of the platform is to provide data about the number of people moving in front of a digital advertisment, and thus enable advertisment providers to estimate their audience volume. Since the system is already at production level, we managed to perform an internal evaluation of its functionality, to prove its effectiveness and create a baseline for further improvements.

In this work we focus only on pedestrians in close range to the advertisement medium - the full audience consists of people in vehicles such as cars and busses, as well as non-pedestrians e.g. people sitting in cafes and restaurants, yet those are out of scope of this evaluation.

1.1 Related Works

The modern ARA system, described in this paper, is based on the ARAHUB platform presented in [7]. The new ARA system was created to improve mostly on the outdoor scenarios. In particular, the imaging sensor was re-designed to provide a wide field of view by using two cameras. ARA sensors were also modified to be standalone devices dedicated for observing a single point of interest.

Similar human audience counting systems were investigated in [1,5,3,2]. Commercial solutions from companies like Quividi or Aquaji exist on the market, however detailed descriptions of their accuracy and evaluation methodology are not publicly available for comparison. A similar embedded system, but for the rescue purposes, was proposed by Gąsienica-Józkowy et al. [4]. Its main component is a weighted YOLOv3 based detector trained with silhouettes of floating persons for rescue reasons. On the other hand, an underwater detection system based on the SSD detector was proposed by Knapik et al. [6].

2 System architecture

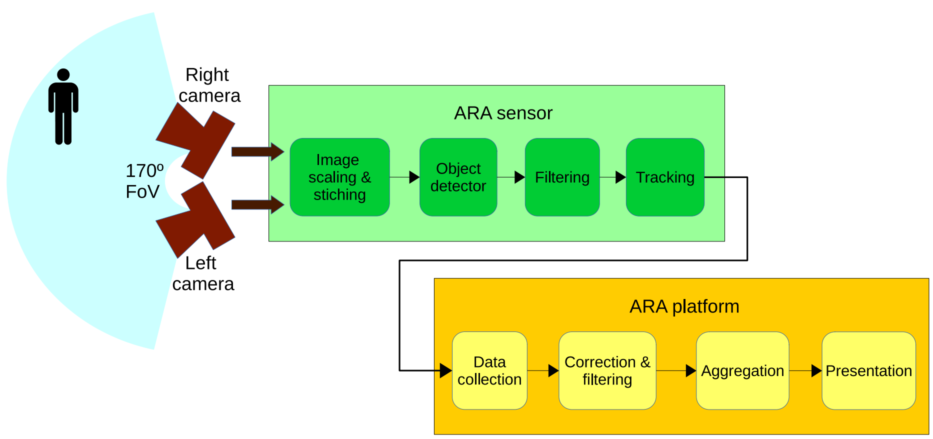

Fig. 1. ARA system architecture: images from two cameras are processed by ARA sensor to detect objects. Movement paths of humans are transmitted to ARA platform for storage, processing and presentation.

The ARA system, presented in Figure 1, consists of two main parts: a) a network of independent embedded devices with GPU acceleration and equipped with a pair of cameras (ARA sensor); b) a central service for data storage, processing and presentation (ARA Platform). The ARA sensor is based on NVidia's jetson nano - an embedded ARM platform equipped with CUDA-capable GPU accelerator, dedicated for use in ML applications. The device includes two HD cameras, which are set side-by-side in order to observe a FoV with the 170° horizontal and 110° vertical angles, respectively.

Images from these two cameras are stitched together using the cylindrical projection to form a single panoramic image. Next, the image is resized to 384x160 pixels and objects are detected using YOLOv4-CSP [8]. The network is retrained on a custom dataset for improved person detection in both day and night lighting conditions.

Objects that do not meet conditions regarding size and certainty are filtered out. A tracking algorithm based on the Kalman filter is applied to objects detected in subsequent frames to estimate their movement paths. Such paths are then transmitted to the ARA platform.

ARA platform receives and stores movement paths from a large number of devices mounted in different localizations. The data is conditioned and filtered to remove paths that are too short, anomalous or fall out of manually defined areas of interest. Objects are then counted and aggregated statistics for various time periods are calculated. Finally, the platform provides web interface for presenting the data in an informative visual form.

3 Evaluation

3.1 Methodology

In order to evaluate the accuracy of the ARA system for pedestrian count we have prepared a custom dataset that consists of fragments of video footage recorded by the ARA sensor. To obtain ground truth values - the real number of pedestrians - the dataset was manually annotated by a group of researchers. For each footage an area of interest was defined - both human annotators and the ML model counted persons passing through that area, neglecting other visible objects. The dataset was split randomly and uniformly between 8 annotators, each annotated both day and night recordings. Next, the same dataset was processed by the model used in the ARA-sensors. Then, detected movement paths were parsed by the ARA-platform and the resulting pedestrian counts were stored.

3.2 Results

The results are presented in Table 1. Recorded samples were taken from 18 locations during the day and night in different weather conditions and different distances from the camera. The density of pedestrians vary in the selected locations. The dataset consisted of 207 minutes of real-time recordings. In total, almost 1000 pedestrians were visible. The total number of pedestrians detected by the system is close to the manually counted value, however the system has noticeable variance. It is worth noticing that we do not know the manual counting error. In locations with perfect observation conditions, the error can be as low as 5% during the day and 7% at night. In a few locations the system was working in almost complete darkness, yet it was capable of counting at least some pedestrians.

Table 1. Results for pedestrian count for 18 locations in both day and night conditions. The total number of pedestrians are presented for manual and automated count

	Locations number	Manual count	Estimated count	Estimated count MAPE		
				Mean	9-th decile	1-st decile
Day conditions	18	534	559	13.8%	5.3%	30.5%
Night conditions	18	407	424	29.8%	7.3%	62.8%

4 Conclusions

In this paper we have shown the evaluation results of the ARA system - in particular its capability to count human audience in the outdoor digital advertisement conditions. The accuracy of the AI system based on DNN, running on an inexpensive embedded system, is comparable to human results in the daylight scenarios and slightly worse in the nighttime conditions. As the system is not universally applicable yet, it may provide useful data when used in the right environment.

Acknowledgments

This work was supported by the National Centre for Research and Development, Poland, under the grant no. POIR.01.01.01-00-1116/20.

References

1. Brunetti, A., Buongiorno, D., Trotta, G.F., Bevilacqua, V.: Computer vision and deep learning techniques for pedestrian detection and tracking: A survey. Neurocomputing **300**, 17–33 (2018)
2. Cyganek, B., Woźniak, M.: Tensor-based shot boundary detection in video streams. New Generation Computing **35**(4), 311–340 (Oct 2017)
3. F. Fleuret, J. Berclaz, R.L., Fua, P.: Multi-camera people tracking with a probabilistic occupancy map. IEEE Transactions on Pattern Analysis and Machine Intelligence **30**(2), 267–282 (2008)
4. Gąsienica-Józkowy, J., Knapik, M., Cyganek, B.: An ensemble deep learning method with optimized weights for drone-based water rescue and surveillance. Integrated Computer-Aided Engineering **28**, 221–235 (2021), 3
5. Iguernaissi, R., Merad, D., Aziz, K., Drap, P.: People tracking in multi-camera systems: a review. Multimedia Tools and Applications **78** (09 2018)
6. Knapik, M., Cyganek, B.: Evaluation of deep learning strategies for underwater object search. In: First Int. Conf. on Societal Automation (SA). pp. 1–6 (2019)
7. Ludziejewski, J., Grad, Ł., Przebinda, Ł., Tajmajer, T.: Integrated human tracking based on video and smartphone signal processing within the ARAHUB system. In: 15th Conference on Computer Science and Information Systems (FedCSIS). pp. 105–114 (2020)
8. Wang, C., Bochkovskiy, A., Liao, H.: Scaled-YOLOv4: Scaling cross stage partial network. In: 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 13024–13033. IEEE Computer Society (2021)