AIMH Lab: Smart Cameras for Public Administration

Luca Ciampi, Donato Cafarelli, Fabio Carrara, Marco Di Benedetto, Fabrizio Falchi, Claudio Gennaro, Fabio Valerio Massoli, Nicola Messina, Claudio Vairo, Giuseppe Amato

Artificial Intelligence for Media and Humanities laboratory
Istituto di Scienza e Tecnologie dell'Informazione "A. Faedo", CNR
<name.surname>@isti.cnr.it

Abstract

In this short paper, we report the activities of the Artificial Intelligence for Media and Humanities (AIMH) laboratory of the ISTI-CNR related to Public Administration. In particular, we present some AI-based public services serving the citizens that help achieve common goals beneficial to the society, putting humans at the epicenter. Through the automatic analysis of images gathered from city cameras, we provide AI applications ranging from smart parking and smart mobility to human activity monitoring.

1 Introduction

New technologies, such as Artificial Intelligence (AI), are playing a significant role in the modernization and overall improvement of the functioning of Public Administration (PA). Today, AI can drive vehicles, personalize shopping, or take care of elderly or sick people, to give a few examples. In the PA, the potential is enormous. For instance, AI can be used profitably in healthcare by exploiting algorithms able to automatically read results of medical exams or in the management of relations with citizens using chatbots to answer people's queries cutting through layers of bureaucracy. In general, AI is increasingly employed to develop public services that are making life easier for citizens.

In this paper, we present some research topics and applications carried out by the Artificial Intelligence for Media and Humanities (AIMH) laboratory of the ISTI-CNR focusing on the study and development of AI-based Public Services dedicated to the analysis and interaction with the physical world can significantly impact human life. These systems can process a massive amount of data and make/suggest decisions that help to solve many real-world problems where *humans are at the epicenter*. Crucial examples of *human-centered AI*, whose aim is to create a better world by achieving common goals beneficial to our societies, are city mobility, pollution monitoring, or critical infrastructure management, where decision-makers require, for instance, measurements about flows of bicycles, cars or people. Specifically, we illustrate some applications based on the analysis of images gathered from city cameras. Like no other sensing mechanism, networks of city cameras can observe the physical world and simultaneously provide visual data to AI systems to extract relevant information from this deluge of data. More in-depth, we discuss some solutions in the sphere of smart mobility that can automatically monitor parking lot occupancy. Moreover, we present an intelligent video surveillance system that can precisely localize the pedestrians present in the monitored scene, that has the peculiarity to have learned the task from images gathered from a video game. Then, we illustrate some applications capable of monitoring vehicle flows in urban scenarios by estimating the traffic density. Furthermore, we describe an AI-based system that can classify human emotions by analyzing facial expressions. Finally, we introduce an AI-assisted framework that can carry out several tasks to help monitor individual and collective human safety rules, such as social distance calculation and facial masks detection.

2 Research Themes and Applications

2.1 Visual Parking Lot Monitoring

Traffic-related issues are constantly increasing, and tomorrow's cities cannot be considered intelligent if they do not enable smart mobility. Nowadays, smart mobility applications, such as smart parking and road traffic management, are widely employed worldwide, making our cities more livable.

City camera networks have become pervasive, and they represent the perfect tool to monitor large areas while simultaneously providing visual data to AI systems in charge of extracting relevant information from this deluge of data. However, this application is often hampered by the limited computational resources on disposable devices. Indeed, Deep Learning (DL)-based solutions, including Convolutional Neural Networks (CNNs), usually require a considerable amount of computational resources, often limiting their applications only on powerful centralized servers.

The AIMH laboratory proposes some DL-based solutions for parking lot monitoring running directly onboard embedded vision systems, i.e., devices equipped with limited computational capabilities that can capture images, process them, and eventually communicate with other devices sending the elaborated information. In particular, in [Amato *et al.*, 2016] and [Amato *et al.*, 2017], we introduce a decentralized and efficient solution for visual parking lot occupancy detection, which exploits CNNs to classify the parking space occupancy. It runs directly onboard smart cameras built using

Raspberry Pi platform equipped with a camera module. On the other hand, in [Ciampi *et al.*, 2018], [Amato *et al.*, 2018] and [Amato *et al.*, 2019a], we extend this application by proposing a DL-based method that is instead able to estimate the number of vehicles present in the Field Of View of the smart cameras. Such a task is more flexible than the previous one since it does not rely on meta-information regarding the monitored scene, such as the position of the parking lots. We show the output of our vehicle counting solution in Fig. 1. Moreover, unlike most of the works on this task, which focuses on the analysis of *single* images, the AIMH group introduces, in [Ciampi *et al.*, 2021a], the use of multiple visual sources to monitor a wider parking area from different perspectives. The proposed multi-camera system is capable of automatically estimating the number of cars present in the *entire* parking lot directly on board the edge devices. It comprises an on-device DL-based detector that locates and counts the vehicles from the captured images and a decentralized geometric-based approach that can analyze the inter-camera shared areas and merge the data acquired by all the devices. Finally, in [Amato *et al.*, 2019b] we propose a DL solution to automatically detect and count vehicles in images taken from drones.

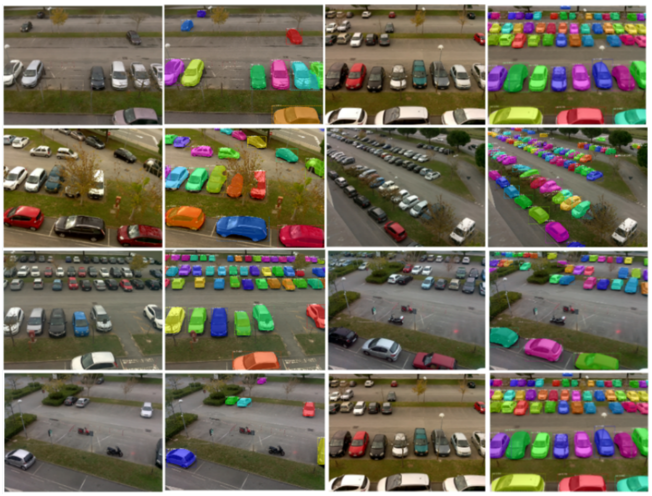

Figure 1: **Examples of the output of our vehicle counting solution.** We show input images captured by smart cameras and the detected vehicles by our CNN-based technique.

2.2 Virtual To Real Adaptation of Pedestrian Detectors

An essential task in many intelligent video surveillance systems is pedestrian detection since it is the main building block for many applications, such as people re-identification. CNN-based pedestrian detectors have demonstrated their superiority over the approaches relying on hand-crafted features. However, the crux of CNNs is that to generalize well at inference time, they require a massive amount of diverse labeled data during the training phase, covering the widest number of different scenarios. Since manually annotating new collections of images is expensive and requires a significant human effort, an appealing solution is to gather synthetic data from virtual environments resembling the real world, where the labels are *automatically* collected interacting with the graphical engine. In this direction, the AIMH group introduces *Virtual Pedestrian Dataset (ViPeD)* [Amato *et al.*, 2019c], a new synthetic dataset generated with the highly photo-realistic graphical engine of a video game. We show a sample of this dataset in Fig. 2. We exploited it to train the CNN-based pedestrian detector. However, data coming from virtual worlds cannot be fully exploited due to the Synthetic-to-Real Domain Shift, i.e., the image appearance difference between the synthetic training data and the real-world ones on which the pedestrian detector, in the end, shall be used. This domain gap between the two data distributions leads to performance degradation of the CNN at test time, and so, intending to mitigate it, we propose two different *Supervised Domain Adaptation* strategies [Ciampi *et al.*, 2020b].

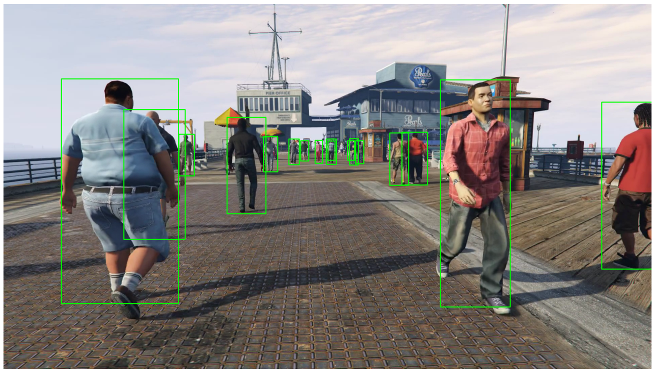

Figure 2: **Sample of our *ViPeD* dataset**. Images and bounding boxes localizing the pedestrians are *automatically* gathered from a virtual-world. Image Courtesy of [Ciampi *et al.*, 2020b].

2.3 Visual Traffic Density Estimation

Monitoring vehicle flows in cities is crucial to improving citizens' urban environment and quality of life, and images are the best sensing modality to perceive and assess the flow of vehicles in large areas. Nowadays, many AI-based systems that analyze this massive amount of visual data coming from city camera networks are emerging. However, these machine learning-based technologies hinge on large quantities of annotated data, preventing their scalability to city-scale as new cameras are added to the system. Scenarios that are never seen during the supervised training phase systematically lead to performance degradation of these approaches due to the existence of a *Domain Shift* between the distributions of the training and test data.

The AIMH laboratory proposes a technique that can automatically estimate the traffic density of urban scenarios by analyzing images [Ciampi *et al.*, 2020c]. The main peculiarity of the proposed methodology is that it can generalize to new sources of data for which there is no training data available. We achieved this generalization by exploiting an *Unsupervised Domain Adaptation (UDA)* strategy, whereby a discriminator attached to the output induces similar density distribution in the test and train domains. Furthermore, we extend this work in [Ciampi *et al.*, 2020a], [Ciampi *et al.*, 2021c] and [Ciampi *et al.*, 2021b], introducing the *Grand Traffic Auto (GTA)* dataset, the first collection of images with precise *per-pixel* annotations gathered using the graphical engine of a video game. We show a sample of this dataset in Fig. 3. Exploiting our UDA methodology, we mitigated the domain gap existing between the synthetic and the real-world images in an *unsupervised* fashion.

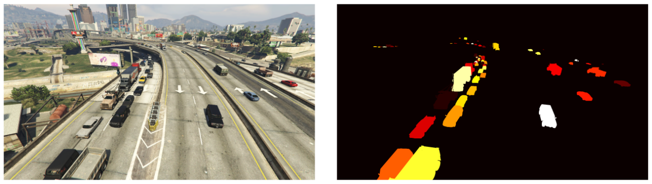

Figure 3: **Sample of our *GTA* dataset.** It is the first *synthetic* dataset of urban scenarios where per-pixel annotations are *automatically* gathered from a virtual-world.

2.4 Facial Expression Recognition

Facial expressions play a fundamental role in human communication. Their study, which represents a multidisciplinary subject, embraces a great variety of research fields, e.g., psychology, computer science, among others. Concerning DL, recognizing facial expressions is a task named Facial Expression Recognition (FER). With such an objective, the goal of a learning model is to classify human emotions starting from a facial image of a given subject. Typically, face images are acquired by cameras that have, by nature, different characteristics, such as the output resolution. Moreover, other circumstances might involve cameras placed far from the observed scene, thus obtaining faces with very low resolutions. Therefore, since the FER task might involve analyzing face images that can be acquired with heterogeneous sources, it is plausible to expect that resolution plays a vital role. In such a context, the AIMH group proposes a multi-resolution training approach to solve the FER task ([Massoli *et al.*, 2021b], [Cafarelli *et al.*, 2021], [Massoli *et al.*, 2021a]). We grounded our intuition on the observation that, often, face images are acquired at different resolutions. Thus, directly considering such property while training a model can help achieve higher performance on recognizing facial expressions. We show in Fig. 4 an example of the output of our solution.

2.5 Human Activity Monitoring

As occurs during a severe health emergency event, there exist scenarios in which ensuring compliance to a set of guidelines becomes crucial to secure a safe living environment in which human activities can be conducted. In fact, as evidenced during the recent COVID-19 pandemic, wearing medical masks, avoiding the creation of large gatherings in confined places, and keeping a certain physical distance among people were the most common rules every government applied in their jurisdiction territories. However, human supervision could not always guarantee this task, especially in crowded scenes where checking usage of personal protection equipment or enforcing strict social behavior has to be continuously assessed to preserve global health. To this end, the AIMH group presents a deployed real use-case embedded system capable of perceiving people's behavior and aggregations, and able to supervise the appliance of a set of rules relying on a configurable plug-in framework [Di Benedetto *et al.*, 2021]. Working on indoor and outdoor environments, we demonstrated that our implementation of counting people aggregations, measuring their reciprocal physical distances, and checking the proper usage of protective equipment is a practical yet open framework for monitoring human activities in critical conditions. In Fig. 5, we show an example of the functionality aiming at estimating the social distance.

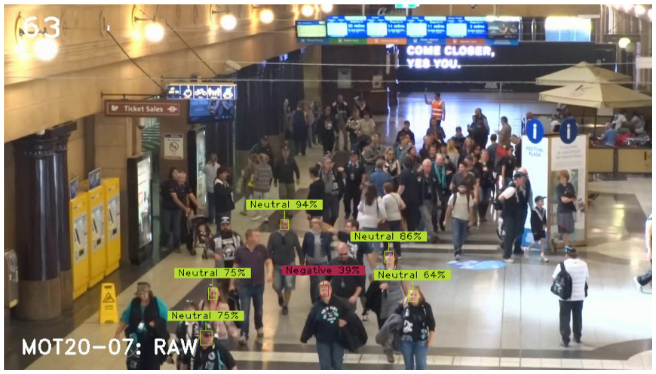

Figure 4: **Example of the output of our Facial Expression Recognition system.** We recognize human facial expressions by automatically analyzing images.

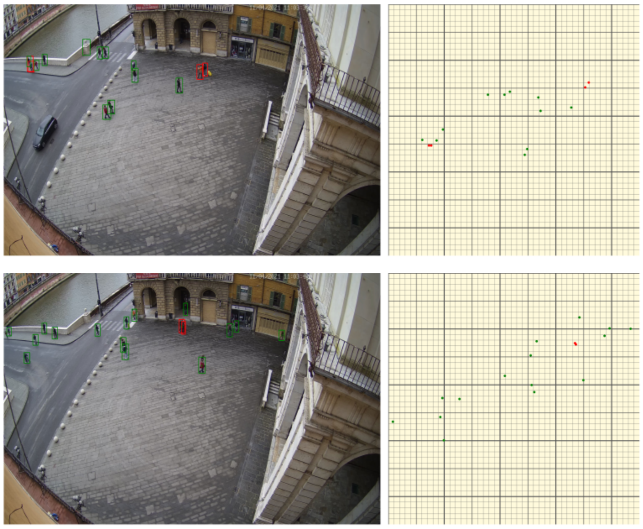

Figure 5: **Examples of the output of our social distance measurer module.** *Left*: images with the detected pedestrians. *Right*: 2D projection on a virtual planar surface obtained through homography. Green color means a safe placement; red color indicates violations of the social distance rule. Image Courtesy of [Di Benedetto *et al.*, 2021].

3 Projects

AI4EU (A European AI On Demand Platform and Ecosystem) – The activities of the H2020 AI4EU project include the design of a European AI on-Demand Platform to support this ecosystem and share AI resources produced in European projects.

CROWDVISOR – It is a technology-tranferred project aimed at human activity monitoring using Computer Vision and AI funded by ARTES 4.0. The system has been embedded on a dedicated low-cost device and deployed as a urban administration monitoring service in the city of Pisa.

WeAreClouds@Lucca – funded by Fondazione Cassa di Risparmio di Lucca, carries out research and development in the field of monitoring public places through cameras and microphones.

AI4CHSites – Artificial Intelligence for monitoring Cultural Heritage Sites co-funded by Tuscany Region (Italy). Prototypes are tested on the Square of Miracles in Pisa including the Leaning Tower.

AI4Media – A Centre of Excellence delivering next generation AI Research and Training at the service of Media, Society and Democracy fundend by EU.

4 Challenges

Many open challenges need to be addressed in the future. We will try to integrate and expand modules of our human activity monitoring framework with further visual analyses, like gesture/posture recognition. We will also attempt to apply a transfer learning approach to predict physical distances among people by using an automatically labeled computer-generated training set based on a rendering engine simulation. Other challenging directions are the creation of additional synthetic datasets suitable for the tracking tasks, i.e., where the instances of the objects are tracked during the time, and the introduction of other Domain Adaptation techniques to fill the gap between real- and virtual-world images.

References

[Amato *et al.*, 2016] G. Amato, F. Carrara, F. Falchi, C. Gennaro, e C. Vairo. Car parking occupancy detection using smart camera networks and deep learning. In *2016 IEEE Symposium on Computers and Communication (ISCC)*, pages 1212–1217, 2016.

[Amato *et al.*, 2017] G. Amato, F. Carrara, F. Falchi, C. Gennaro, C. Meghini, e C. Vairo. Deep learning for decentralized parking lot occupancy detection. *Expert Systems with Applications*, 72:327–334, 2017.

[Amato *et al.*, 2018] G. Amato, P. Bolettieri, D. Moroni, F. Carrara, L. Ciampi, G. Pieri, C. Gennaro, G. R. Leone, e C. Vairo. A wireless smart camera network for parking monitoring. In *2018 IEEE Globecom Workshops*, dec 2018.

[Amato *et al.*, 2019a] G. Amato, P. Bolettieri, F. Carrara, L. Ciampi, C. Gennaro, G. R. Leone, D. Moroni, G. Pieri, e C. Vairo. Parking lot monitoring with smart cameras. In *5th Italian Conference on ICT for Smart Cities And Communities*, 2019.

[Amato *et al.*, 2019b] G. Amato, L. Ciampi, F. Falchi, e C. Gennaro. Counting vehicles with deep learning in onboard UAV imagery. In *2019 IEEE Symposium on Computers and Communications (ISCC)*. IEEE, jun 2019.

[Amato *et al.*, 2019c] G. Amato, L. Ciampi, F. Falchi, C. Gennaro, e N. Messina. Learning pedestrian detection from virtual worlds. In *International Conference on Image Analysis and Processing (ICIAP)*, pages 302–312. Springer, 2019.

[Cafarelli *et al.*, 2021] D. Cafarelli, F. V. Massoli, F. Falchi, C. Gennaro, e G. Amato. Expression recognition analysis in the wild. *arXiv:2101.09231*, 2021.

[Ciampi *et al.*, 2018] L. Ciampi, G. Amato, F. Falchi, C. Gennaro, e F. Rabitti. Counting vehicles with cameras. In *Proceedings of the 26th Italian Symposium on Advanced Database Systems*, volume 2161 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2018.

[Ciampi *et al.*, 2020a] L. Ciampi, C. Gennaro, e G. Amato. Monitoring traffic flows via unsupervised domain adaptation. In *6th Italian Conference on ICT for Smart Cities And Communities*, 2020.

[Ciampi *et al.*, 2020b] L. Ciampi, N. Messina, F. Falchi, C. Gennaro, e G. Amato. Virtual to real adaptation of pedestrian detectors. *Sensors*, 20(18):5250, sep 2020.

[Ciampi *et al.*, 2020c] L. Ciampi, C. Santiago, J. P. Costeira, C. Gennaro, e G. Amato. Unsupervised vehicle counting via multiple camera domain adaptation. In *Proceedings of the First International Workshop on New Foundations for Human-Centered AI (NeHuAI)*, pages 82–85, 2020.

[Ciampi *et al.*, 2021a] L. Ciampi, C. Gennaro, F. Carrara, F. Falchi, C. Vairo, e G. Amato. Multi-camera vehicle counting using edge-ai. *arXiv:2106.02842*, 2021.

[Ciampi *et al.*, 2021b] L. Ciampi, C. Santiago, J. Costeira, C. Gennaro, e G. Amato. Domain adaptation for traffic density estimation. In *Proceedings of the 16th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*, 2021.

[Ciampi *et al.*, 2021c] L. Ciampi, C. Santiago, J. P. Costeira, C. Gennaro, e G. Amato. Traffic density estimation via unsupervised domain adaptation (discussion paper). In *Proceedings of the 29th Italian Symposium on Advanced Database Systems, SEBD 2021*, pages 442–449, 2021.

[Di Benedetto *et al.*, 2021] M. Di Benedetto, F. Carrara, L. Ciampi, F. Falchi, C. Gennaro, e G. Amato. An embedded toolset for human activity monitoring in critical environments. *Under Review.*, 2021.

[Massoli *et al.*, 2021a] F. V. Massoli, D. Cafarelli, G. Amato, e F. Falchi. A multi-resolution training for expression recognition in the wild. In *SEBD 2021 - Italian Symposium on Advanced Database Systems*, pages 427–433, 2021.

[Massoli *et al.*, 2021b] F. V. Massoli, D. Cafarelli, C. Gennaro, G. Amato, e F. Falchi. MAFER: a multi-resolution approach to facial expression recognition. *arXiv:2105.02481*, 2021.