MobiNetVideo Newsletters

#1 - September 2018

TEC2017-88169-R MobiNetVideo (2018-2020)

Visual Analysis for Practical Deployment of Cooperative Mobile Camera Networks

http://www-vpu.eps.uam.es/MobiNetVideo/

Introducing MobiNetVideo

The main objective of this project is to achieve advances in the state-of-the-art for video analysis systems employing mobile imaging devices, as a response to the growing social and industry demand for potential visual-based applications supported by the boom of mobile devices featuring high imaging and computing capabilities.

This project will focus on networked systems composed of fixed, pan-tilt-zoom and mobile cameras (i.e., heterogeneous camera networks) and it will address key challenges that nowadays limit their widespread adoption: autonomous and coordinated operation, performance of visual analysis tools and practical deployment issues. The outcomes of this project will underpin emerging applications in many areas where camera mobility plays a key role, such as unmanned aerial vehicles exploration for search and rescue missions and disaster management, smart traffic and people surveillance, in-home pervasive networks for monitoring the elderly, real-time health supervision and life-logging cameras integrated with wearable body sensors, etc.

Based on our experience on the topic, this project starts from the hypothesis that most of the operating video analysis systems rely on quite constrained infrastructures mainly characterized by static and pan-tilt-zoom cameras. These sets of cameras are not usually managed as networked devices since cameras run independently, mainly for storage operations. Such systems also present very limited mobility which restricts their ability for adapting to unforeseen conditions while concurrently performing various tasks. Moreover, many existing approaches for camera networks focus on a theoretical basis, just being validated by few controlled laboratory experiments. Therefore, practical deployment issues such as robustness to network constraints, resource availability, scalability, etc., are not often considered. All these abovementioned limitations prevent the use of current approaches in practical scenarios.

This context brings new research opportunities, in innovative approaches that can take advantage of the capabilities provided by networked mobile cameras. This project focuses its efforts on three related aspects: 1) visual analysis, 2) cooperation in camera networks and 3) practical deployment.

- The goals for visual analysis aim to achieve a space-time description of the visual content captured by a single mobile camera.
- The cooperation will explore decentralized and distributed schemes to allocate tasks during runtime via task-related decisions (e.g., learning the appearance of moving targets) and operational decisions (e.g., information sharing across cameras).
- Practical deployment will deepen into self-adaptation to manage the available resources (e.g., sensing and computation), handling of the operational constraints (e.g., real-time and communication) and an efficient use of the application environment (e.g., heat maps for detection).

The research achievements will be put in practice for selected case studies reflecting the potential applications of mobile camera networks. To overcome the limited availability of real-world data for complex heterogeneous camera networks, this project considers simulation as an intermediate stage to validate performance objectives before deployment. Virtual reality simulation will be jointly used with communication and hardware simulation to provide a rich test environment where approaches can be applied over situations that can be repeated as needed.

First nine months progress report

The provisional announcement of project approval arrived in the first days of 2018, but the official project approval was not published till mid-June, allowing to start the official opening of the project at the university. Whilst waiting for the notification of definitive approval, we have been working in the lines described in the proposal, but without being able to work with financial support.

After launching the project web page on June 14th, during July we have been adjusting the workplan to the approved budget, adjustment that yielded to the delay of D2 and D5v1, as well as the delay of the first Newsletter (this one). During September, we have completed some adjustments in the workplan, that are detailed below.

Within WP1 we have delayed milestones M1.1 and M1.2. WP2 tasks have been extended till M15 and all milestones and deliverables have been scheduled for the end of WP activities. WP3 and WP4 schedules are not modified and we will do our best to work in parallel with WP2 conclusion and WP3 launching. With respect to WP5, the first developers' workshop has been cancelled and D5v1 has been rescheduled to M12. Therefore, during this first nine months we have been working mainly in the tasks of WP2:

- T2.1 People tracking for active vision: the work has focused on: 1) Adapting the existing tracking system to a portable setting in order to ease data-sets generation and testing; 2) Analyzing existing PTZ cameras to find models with a minimum time-response to motion commands; 3) Evaluating the adequation of detector-based trackers for this kind of real-time applications. Currently, we are documenting the achieved conclusions and focusing on the tracker/camera limitations in these scenarios.
- T2.2 Object detection for collision detection: the work for this task has been mainly focused on evaluating state-of-the-art approaches for multiple class object detectors in data captured from a single camera on-board of an aerial unmanned vehicle. In the next months, we will document these evaluations and explore the use of other public datasets.
- T2.3 Scene categorization for life-logging: the work has focused mainly on scene recognition and semantic segmentation by performing a comparative evaluation of scene recognition methods based on convolutional neural networks, both in overall and in scene-dependent terms, and designing and developing a mechanism which, based on net attention maps and semantic information, is able to *refocus* the net, i.e. to change the image areas used for the recognition. Our current goal is to explore how to use semantic information earlier in the process of scene recognition.
- T2.4 Multi-target tracking for UAV monitoring: the work has been mainly focused on evaluating state-of-the-art approaches for multiple object tracking (MOT) in data captured from a single camera on-board of an aerial unmanned vehicle. In particular, three contributions have been achieved: 1) Development of a framework for MOT in Python which employs spatial and appearance (deep) features, 2) Implementation of evaluation protocols to estimate tracking performance and 3) Participation in the VisDrone challenge 2018 (ECCV workshop) http://aiskyeye.com/. We will further document these contributions and explore the use of other public datasets as well as the relationship between detection and tracking.

First nine months results: Master Thesis

Master thesis

Mejora del rendimiento de redes convolucionales entrenadas para el reconocimiento de escena mediante el uso de información sobre los objetos comunes a éstas (**Refocusing a scene recognition ConvNet by incorporating object priors**), Raúl García Jiménez, (advisor: Marcos Escudero Viñolo), Trabajo Fin de Máster (Master Thesis), Master en Ingeniería de Telecomunicación, Univ. Autónoma de Madrid, Jul. 2018.

Abstract: This work studies the performance of schemes based on convolutional networks in the task of automatic scenes recognition. The work begins with a brief qualitative study of five of the most used architectures. Subsequently, the performance of these architectures is evaluated quantitatively in the task of scene recognition, particularizing in the effect on the performance of the category and the nets architecture. Likewise, the responses of the networks to images that represent untrained scene categories are studied. Furthermore, the robustness of the analyzed solutions to image noise is also evaluated. The results of this quantitative study motivate the design and development of a scheme to improve the performance of networks without having to retrain or adjust them. For this purpose, a refocusing mechanism of one of the studied convolutional networks is outlined. This scheme makes use of a complementary neural network that is trained from descriptions of the scene based on the objects present in it. To obtain these descriptions, another convolutional network that is designed and trained to obtain the semantic segmentation of an image is used. This network provides object-wise annotations of each image pixel. The annotations thereby obtained are weighted by a focusing information of the scene convolutional network, giving more relevance to the areas of the image with the greatest impact on prediction. The predictions of the convolutional network and this scene neural network are compared in an iterative consensus scheme. In this scheme, the image is gradually modified if the predictions of both networks do not coincide, forcing the convolutional network to rely on different areas of the image to predict the scene. Preliminary results in a subset of the analyzed database are promising, reaching relative improvements of 8.85% with respect to the performance of the scene convolutional network.

The MobiNetVideo (TEC2017-88169-R) project is supported by

GOBIERNO DE ESPAÑA MINISTERIO DE ECONOMÍA, INDUSTRIA Y COMPETITIVIDAD