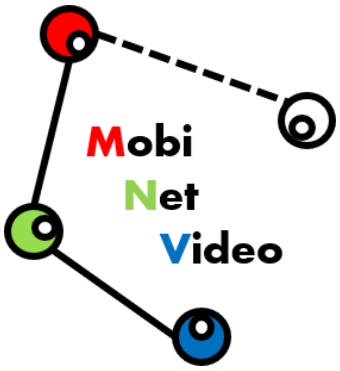

TEC2017-88169-R MobiNetVideo (2018-2020-2021)

Visual Analysis for Practical Deployment of Cooperative Mobile Camera Networks

D5 v4

Results Report

Video Processing and Understanding Lab

Escuela Politécnica Superior

Universidad Autónoma de Madrid

Supported by

AUTHORS LIST

José M. Martínez	josem.martinez@uam.es
Pablo Carballeira López	Pablo.carballeira@uam.es
Juan Carlos San Miguel Avedillo	juancarlos.sanmiguel@uam.es
Marcos Escudero Viñolo	marcos.escudero@uam.es
Álvaro García-Martín	alvaro.garcia@uam.es
Jesús Bescós Cano	j.bescos@uam.es
Miguel Ángel García	miguelangel.garcia@uam.es

HISTORY

Version	Date	Editor	Description
0.1	06/06/2019	José M. Martínez	First Draft version
0.2	12/06/2019	José M. Martínez	Structure for Contributions and draft for several sections
0.3		Pablo Carballeira	Contributions
0.4		Juan Carlos San Miguel	Contributions
0.5		Marcos Escudero	Contributions
0.6		Álvaro García-Martín	Contributions
0.7		José M. Martínez	Working Draft edition
0.7		Marcos Escudero	Working Draft comments
0.8		Pablo Carballeira	Working Draft comments
0.9		Álvaro García-Martín	Working Draft comments
0.10		Juan Carlos San Miguel	Working Draft comments
0.11	31/07/2019	José M. Martínez	Final Working Draft – Editorial checking
1.0	31/07/2019		First version
1.1	10/03/2020	José M. Martínez	Structure for Contributions and draft for several sections
1.2	13/03/2020	Marcos Escudero	Contributions
1.3	15/03/2020	Juan Carlos San Miguel	Contributions

1.4	20/03/2020	Álvaro García-Martín	Contributions
1.5	22/03/2020	Pablo Carballeira López	Contributions
1.5	24/03/2020	Jesús Bescós Cano	Contributions
1.6	30/03/2020	José M. Martínez	Final Working Draft – Editorial checking
2.0	30/07/2020		Second version
2.1	04/11/2020	José M. Martínez	Structure for Contributions and draft for several sections
2.4	09/12/2020	Álvaro García-Martín	Contributions
2.3	14/12/2020	Juan Carlos San Miguel	Contributions
2.4	18/12/2020	Marcos Escudero	Contributions
2.5	21/12/2020	Pablo Carballeira López	Contributions
2.5	22/12/2020	Miguel Ángel García García	Contributions
2.6	23/12/2020	Jesús Bescós Cano	Contributions
2.7	28/12/2020	José M. Martínez	Final Working Draft – Editorial checking
3.0	28/12/2020		Third version
3.1	16/07/2021	José M. Martínez	Structure for Contributions
3.2	07/09/2021	Álvaro García-Martín	Contributions
3.3	10/09/2021	Marcos Escudero	Contributions
3.4	12/09/2021	Miguel Ángel García García	Contributions
3.5	17/09/2021	Pablo Carballeria	Contributions
3.6	19/09/2021	Álvaro García-Martín	Contributions
3.7	24/09/2021	Juan Carlos San Migeul	Contributions
3.8	26/09/2021	José M. Martínez	Final Working Draft – Editorial checking
4.0	28/09/2021	José M. Martínez	Fourth version

CONTENTS:

1. Introduction

This *report* recapitulates the results obtained within the MobiNetVideo project. The results are announced as they are obtained at the Web site (http://www-vpu.eps.uam.es/MobiNetVideo).

1.1. Document structure

This document contains the following chapters:

- Chapter 1: Introduction to this document
- Chapter 2: Publications
- Chapter 3: Project Documents
- Chapter 4: Public Resources: Content Sets and Software
- Chapter 5: Workshops and Seminars
- Chapter 6: Main achievements of the project

2. Publications

2.1. Journals

[1] Alejandro López-Cifuentes, Marcos Escudero-Viñolo, Jesús Bescós, Álvaro García-Martín, "Semantic-aware scene recognition, Pattern Recognition", Volume 102, June 2020, 107256, ISSN 0031-3203, (DOI 10.1016/j.patcog.2020.107256)

2.2. Book Chapters

2.3. Conferences

[2] Zhu P. et al., "VisDrone-VDT2018: The Vision Meets Drone Video Detection and Tracking Challenge Results", in Computer Vision – ECCV 2018 Workshops. ECCV 2018, L. Leal-Taixé, S. Roth (eds.) Lecture Notes in Computer Science, Vol. 11133. Springer, Cham, 2019, pp. 496-518. (ISBN 978-3-030-11020-8) (DOI 10.1007/978-3-030-11021-5_29)

[3] Elena Luna, Paula Moral, Juan C. SanMiguel, Álvaro García-Martín, José M. Martínez, "VPULab participation at AI City Challenge 2019", Proc. of IEEE Int. Conf. on Computer Vision and Pattern Recognition Workshops (CVPRW2019), Long Beach, CA, USA, Jun. 2019, in press. (CVF Open Access version)

[4] Paula Moral, Álvaro García-Martín, José M. Martínez, "Vehicle Re-Identification in Multi-Camera scenarios based on Ensembling Deep Learning Features", Proc. of IEEE Int. Conf. on Computer Vision and Pattern Recognition Workshops (CVPRW2020), Seattle, Washington, USA, Jun. 2020 (DOI 10.1109/CVPRW50498.2020.00310).

[5] Alejandro López-Cifuentes, Marcos Escudero-Viñolo, Jesús Bescós, "A Prospective Study on Sequence-Driven Temporal Sampling and Ego-Motion Compensation for Action Recognition in the EPIC-Kitchens Dataset", Proc. of IEEE Int. Conf. on Computer Vision and Pattern Recognition Workshops (CVPR2020), Seattle, Washington, USA, Jan. 2021. (arXiv:2008.11588)

[6] M. Kristan et al., "The Eighth Visual Object Tracking VOT2020 Challenge Results", ECCVW 2020, Onñine, Aug. 2020, in press

[7] Marta Fernández, Paula Moral, Álvaro García-Martín and José M. Martínez, "Vehicle Re-Identification based on Ensembling Deep Learning Features including a Synthetic Training Dataset, Orientation and Background Features, and Camera Verification.", AI City Challenge Workshop, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, 2021, pp. 4068-4076. (CVF Open Access version).

[8] A. López-Cifuentes, Marcos Escudero-Viñolo, Andrija Gajić, Jesús Bescós, "Visualizing the Effect of Semantic Classes in the Attribution of Scene Recognition Models" in Workshop "EXPLAINABLE DEEP LEARNING- AI" of the 25th International Conference on Pattern Recognition (ICPR) Milano, Italy (Jan 10-15, 2021). DOI: 10.1007/978-3-030-68796-0_9.

2.4. PhD Thesis

2.5. Master Thesis

[9] 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.

[10] Object detection and association in multiview scenarios based on Deep Learning, Paula Moral de Eusebio (advisor: Álvaro García-Martín), Trabajo Fin de Máster (Master Thesis), Master en Investigación e Innovación en TIC – Programa Internacional de Múltiple Titulación IPCV (Image Processing and Computer Vision Master Program), Univ. Autónoma de Madrid, Jul. 2019.

[11] Learning how to modify training rates in scene-recognition, Miguel Basarte Mena, (advisor: Marcos Escudero Viñolo), Trabajo Fin de Máster (Master Thesis), Master en Investigación e Innovación en TIC – Programa Internacional de Múltiple Titulación IPCV (Image Processing and Computer Vision), Univ. Autónoma de Madrid, Sep. 2019.

[12] Real-Time Target Tracking to Position a Mobile Device, Awet H. Gebreiwot (advisors: Jesús Bescós, Álvaro García-Martín), Master Thesis, Erasmus Mundus Joint Master Degree in Image Processing and Computer Vision, Univ. Autónoma de Madrid, Jun. 2020.

[13] People detection in omnidirectional cameras: development of a deep learning architecture based on a spatial grid of classifiers, Enrique Sepúlveda Jorcano (advisor: Pablo Carballeira López), Master Thesis, Erasmus Mundus Joint Master Degree in Image Processing and Computer Vision, Univ. Autónoma de Madrid, Jun. 2020.

[14] Desarrollo de aplicaciones móviles de clasificación y detección de objetos a partir de redes convolucionales ligeras (Development of mobile application for object classification and detection based on light convolutional networks), Paulo C. Casa Robles (advisor: Pablo Carballeira López), Trabajo Fin de Máster (Master Thesis), Master en Ingeniería de Telecomunicación, Univ. Autónoma de Madrid, Jun. 2020.

[15] Incorporating Depth in Egocentric Perception, Andrija Gajic (advisor: Marcos Escudero Viñolo), Master Thesis, Erasmus Mundus Joint Master Degree in Image Processing and Computer Vision (IPCV), Univ. Autónoma de Madrid, Jul. 2020.

[16] Análisis automático de vídeo simulado con sistemas multicámaras basados en UNITY (Automatic analysis of video over simulaed multicamera systems based in UNITY), Vinicio Pazmiño Moya (advisor: Juan Carlos San Miguel Avedillo), Trabajo Fin de Máster (Master Thesis), Master en Ingeniería de Telecomunicación, Univ. Autónoma de Madrid, Sep. 2020.

[17] Semantic segmentation in 2D videogames, Javier Montalvo Rodrigo (advisor: Alvaro García Martín), Trabajo Fin de Máster (Master Thesis), Máster

Universitario en Deep Learning for Audio and Video Signal Processing, Univ. Autónoma de Madrid, Jun. 2021.

[18] Application for the demonstration of the automatic registration of transited spaces for contact tracing of infectious diseases using video signals from life-logging cameras, Daniel de Alcalá Valcárcel (advisor: Marcos Escudero-Viñolo), Trabajo Fin de Máster (Master Thesis), Máster Universitario en Deep Learning for Audio and Video Signal Processing, Univ. Autónoma de Madrid, Jun. 2021.

[19] Contributions to the re-identification of objects, Marta Fernández de Barrio (advisor: Alvaro García Martín), Trabajo Fin de Máster (Master Thesis), Máster Universitario en Image Processing and Computer Vision, Univ. Autónoma de Madrid, Jul. 2021.

[20] Aprendizaje y corrección de errores en sistemas de seguimiento basados en redes convolucionales siamesas, Álvaro Iglesias Arias (advisor: Marcos Escudero-Viñolo). Trabajo Fin de Máster (Master Thesis), Master en Ingeniería de Telecomunicación, Univ. Autónoma de Madrid, Jul. 2021.

[21] On the use of synthetic data for semantic segmentation in videos, Roberto Alcober Couso (advisor: Juan Carlos San Miguel), Trabajo Fin de Máster (Master Thesis), Máster Universitario en Deep Learning for Audio and Video Signal Processing, Univ. Autónoma de Madrid, Jul. 2021.

[22] A unified dataset to expan the scope of semantic segmentation. Ana Martín Doncel (advisor: Marcos Escudero-Viñolo). Trabajo Fin de Máster (Master Thesis), Master en Ingeniería de Telecomunicación, Univ. Autónoma de Madrid, Sep. 2021.

[23] Self-paced curriculum learning to cope with unbalanced semantic segmentation datasets. Ana Martín Doncel (advisor: Marcos Escudero-Viñolo). Trabajo Fin de Máster (Master Thesis), Master en Investigación e Innovación en TIC, Univ. Autónoma de Madrid, Sep. 2021.

[24] Evaluation of deep learning-based classification and object detection algorithms for event cameras, Francisco Javier Martín Ameneiro, (advisor: Pablo Carballeira López), Trabajo Fin Máster (Graduate Thesis), Máster Universitario en Deep Learning for Audio and Video Signal Processing, Univ. Autónoma de Madrid, Jul. 2021.

[25] Continual learning for object detection, Carlos Jim enez Mu~noz, (advisor: Alvaro García Martín), Trabajo Fin de Máster (Master Thesis), Máster Universitario en Deep Learning for Audio and Video Signal Processing, Univ. Autónoma de Madrid, Sep. 2021.

[26] Detection-aware multi-object tracking evaluation, Jorge Muñoz Aguado (advisor: Juan Carlos San Miguel), Trabajo Fin de Máster (Master Thesis), Máster Universitario en Deep Learning for Audio and Video Signal Processing, Univ. Autónoma de Madrid, Sep. 2021.

2.6. Graduate Thesis

[27] Detección de objetos en imágenes urbanas de Google Street View (Object detection in urban images from Google Street View), Paula Guerra Toni (advisor: Pablo Carballeira), Trabajo Fin de Grado (Graduate Thesis), Grado en Ingeniería de Tecnologías y Servicios de Telecomunicación, Univ. Autónoma de Madrid, Jul. 2019.

[28] Adaptación de un sistema de detección de personas en cámaras omnidireccionales a descriptores Deep Learning (Adaptation of a system for people detection with omnidirectional cameras to Deep Learning descriptors), Nicolás García Crespo (advisor: Pablo Carballeira), Trabajo Fin de Grado (Graduate Thesis), Grado en Ingeniería de Tecnologías y Servicios de Telecomunicación, Univ. Autónoma de Madrid, Jul. 2019.

[29] Detección jerárquica de grupos de personas con CNNs (Hierarchical detection of groups of people using CNNs), Antonio Campoy Cordero (advisor: Álvaro García-Martín), Trabajo Fin de Grado (Graduate Thesis), Grado en Ingeniería de Tecnologías y Servicios de Telecomunicación, Univ. Autónoma de Madrid, Jul. 2019.

[30] Re-identificación de personas (People re-identification), Daniel Sáez García (advisor: Álvaro García-Martín), Trabajo Fin de Grado (Graduate Thesis), Grado en Ingeniería de Tecnologías y Servicios de Telecomunicación, Univ. Autónoma de Madrid, Jul. 2019.

[31] Reconocimiento de escenas exteriores mediante redes neuronales profundas entrenadas con la base de datos Places (Scene recognition using Deep Neural Networks trained with the Places database), Santiago Vicente Moñivar (advisor: Miguel Ángel García), Trabajo Fin de Grado (Graduate Thesis), Grado en Ingeniería de Tecnologías y Servicios de Telecomunicación, Univ. Autónoma de Madrid, Oct. 2019.

[32] Clasificación de imágenes con redes neuronales profundas mediante conjuntos de entrenamiento reducidos y aprendizaje "few-shot" (Image classification with deep neural networks through reduced training sets and few-shot learning), Guillermo Eliseo Torres Alonso (advisor: Miguel Ángel García), Trabajo Fin de Grado (Graduate Thesis), Grado en Ingeniería de Tecnologías y Servicios de Telecomunicación, Univ. Autónoma de Madrid, Jun. 2020.

[33] Reconocimiento no-supervisado de escenas mediante características extraídas de redes neuronales pre-entrenadas (Unsupervised scene recognition using features extracted from pre-trained neural networks), Alejandro Gilabert Ramírez (advisor: Miguel Ángel García), Trabajo Fin de Grado (Graduate Thesis), Grado en Ingeniería de Tecnologías y Servicios de Telecomunicación, Univ. Autónoma de Madrid, Jul. 2020.

[34] Análisis automático de vídeo simulado con sistemas multi-cámara basados en UNITY (Automatic analysis of video simulated with multi-camera systems based on UNITY), Fernando Terry Sanz-Pastor (advisor: Juan Carlos San Miguel), Trabajo Fin de Grado (Graduate Thesis), Grado en Ingeniería de Tecnologías y Servicios de Telecomunicación, Univ. Autónoma de Madrid, Jul. 2020.

[35] Sistema de Virtualización y gestión de recursos compartidos en un grupo de investigación universitario (System for the virtualization and managment of sahred resoruces in a univerity research group), Manuel Jiménez Sánchez (advisor: Juan Carlos San Miguel), Trabajo Fin de Grado (Graduate Thesis), Grado en Ingeniería Informática, Univ. Autónoma de Madrid, Jul. 2020.

[36] Agrupamiento Espacio-Temporal de Secuencias de Video Mediante Caracterización por la respuesta de Redes Convolucionales (Spatio-temporal grouping of video sequences characterized by convolutional networks response), Julio Moreno Blanco (advisor: Marcos Escudero Viñolo), Trabajo Fin de Grado (Graduate Thesis), Grado en Ingeniería de Tecnologías y Servicios de Telecomunicación, Univ. Autónoma de Madrid, Jul. 2020.

[37] Re-identificación de personas y vehículos (People and vehicles re-identification), Carlos González Ruiz (advisor: Álvaro García-Martín), Trabajo Fin de Grado (Graduate Thesis), Grado en Ingeniería de Tecnologías y Servicios de Telecomunicación, Univ. Autónoma de Madrid, Jul. 2020.

[38] Desarrollo de un marco de trabajo para segmentación semántica en bases de datos de imágenes urbanas (Developmentof a framework for semantic segmentation in urban-image databases), Javier González Cabrero (advisor: Pablo Carballeira), Trabajo Fin de Grado (Graduate Thesis), Grado en Ingeniería de Tecnologías y Servicios de Telecomunicación, Univ. Autónoma de Madrid, Jun. 2020.

[39] Aprendizaje continuo mediante redes convolucionales, Anselmo Velázquez Pazos (advisor: Álvaro García Martín), Trabajo Fin de Grado (Graduate Thesis), Grado en Ingeniería de Tecnologías y Servicios de Telecomunicación, Univ. Autónoma de Madrid, Feb. 2021.

[40] Segmentación objeto-fondo mediante redes convolucionales, Raúl Arcos Serrano (advisor: Álvaro García Martín), Trabajo Fin de Grado (Graduate Thesis), Grado en Ingeniería de Tecnologías y Servicios de Telecomunicación, Univ. Autónoma de Madrid, Jun. 2021.

[41] Detección de colisiones mediante procesado de vídeo, Sergio Avello Largo (advisor: Álvaro García Martín), Trabajo Fin de Grado (Graduate Thesis), Grado en Ingeniería de Tecnologías y Servicios de Telecomunicación, Univ. Autónoma de Madrid, Jul. 2021.

[42] Adaptación de algoritmos de acuerdo con información contextual extraída automáticamente, Javier Santos Gimeno (advisor: Marcos Escudero-Viñolo), Trabajo Fin de Grado (Graduate Thesis), Grado en Ingeniería de Tecnologías y Servicios de Telecomunicación, Univ. Autónoma de Madrid, Jul. 2021.

[43] Ejecución de redes neuronales en móviles Android con aceleración hardware mediante Keras y TensorFlow Lite (Execution of neural networks on Android smartphones with hardware acceleration through Keras and TensorFlow Lite), Ángel Fragua Baeza (advisor: Miguel Ángel García), Trabajo Fin de Grado (Graduate Thesis), Grado en Ingeniería Informática, Univ. Autónoma de Madrid, Jun. 2021.

[44] Reconocimiento de escenas mediante integración multiescala de redes convolucionales (Scene recognition through multi-scale integration of convolutional networks), Pablo Collado Recio (advisor: Miguel Ángel García

García), Trabajo Fin de Grado (Graduate Thesis), Grado en Ingeniería de Tecnologías y Servicios de Telecomunicación, Univ. Autónoma de Madrid, Jun. 2021.

[45] Contribuciones a un sistema de detección de personas en cámaras omnidireccionales, Javier Expósito Cáceres, (advisor: Pablo Carballeira López), Trabajo Fin de Grado (Graduate Thesis), Grado en Ingeniería de Tecnologías y Servicios de Telecomunicación, Univ. Autónoma de Madrid, Jul. 2021.

[46] Desarrollo de un marco de trabajo para la reconstrucción de imágenes a partir de cámaras de eventos, Miguel Lorente Peinado, (advisor: Pablo Carballeira López), Trabajo Fin de Grado (Graduate Thesis), Grado en Ingeniería de Tecnologías y Servicios de Telecomunicación, Univ. Autónoma de Madrid, Jul. 2021

[47] Contribuciones a la simulación de sistemas multi-cámara basados en UNITY (Contributions to the simulation of multi-camera systems based on UNITY), Alejandro Palencia Palomo (advisor: Juan Carlos San Miguel), Trabajo Fin de Grado (Graduate Thesis), Grado en Ingeniería de Tecnologías y Servicios de Telecomunicación, Univ. Autónoma de Madrid, Jul. 2021.

3. Project Documents

3.1. Published Deliverables

3.1.1. D1.1 “System Infrastructure” version 1 (December 2018 - delayed July 2019)

This deliverable describes the infrastructure available at VPULab for the development of the tasks proposed within the MobiNetVideo project (TEC2017-88169-R. Describing both the available hardware for video sequences adquisition (“Adquisition infrastructure”) and the hardware and software environments for data processing and analysis (“Processing infrastructure”).

3.1.2. D1.2 “Camera simulation” version 1 (December 2018 - delayed July 2019)

This deliverable describes the work related with the task T.1.2 Cameras network simulation which supports other tasks for generating test data. We focus on the simulator “Multi-camera System Simulator (MSS)” to describe its structure and the developed features within the context of this project. Moreover, we also integrated other developments for the MSS simulator.

3.1.3. D1.3 “Evaluation datasets” version 1 (December 2018 - delayed July 2019)

This deliverable describes the work related with the task T.1.3 “Generation of datasets. Support to other tasks for generating test data and defining evaluation methodologies”. It includes the selection of appropriate datasets (sequences and associated ground-truth) and the ones generated within the project. Additionally, the evaluation methodologies used for the evaluation for the project algorithms and systems is also described.

3.1.4. D2 “Feasibility studies: algorithms and findings” ((September 2018 - rescheduled March 2019 - delayed July 2019)

This deliverable describes the work related with tasks T.2.1 People tracking for active vision, T.2.2 Object detection for collision detection, T.2.3 Scene categorization for lifelogging and T.2.4 Multi-target tracking for UAV monitoring.

All these WP2 tasks aim at performing a study of current technologies for applications related to heterogeneous camera networks where camera mobility plays a key role. Such studies are performed on public datasets. The main objective is the identification of suitable state-of-the-art video analysis tools (e.g., segmentation, tracking and detection),

by the implementation and evaluation of their performance in single mobile cameras for use as a baseline for comparison with the achievements to be developed within WP3 and WP4.

The technologies covered by the deliverable are:

- People tracking for active vision
- Object detection for collision detection
- Scene categorization for lifelogging
- Multi-target tracking for UAV monitoring

3.1.5. D3 "Technologies for mobile camera networks" version 1 (June 2019 – rescheduled July 2019 – rescheduled September 2019)

This deliverable was rescheduled to September 2019 in July.

This deliverable will describe the new technologies developed within WP3 for applications related to heterogeneous camera networks where camera mobility plays a key role, covering different areas grouped in WP3 tasks:

- T.3.1 Scene Recognition
- T3.2 Semantic Segmentation
- T3.3 Multi-view matching
- T.3.4 Cooperative detection and tracking

3.1.6. D5 "Results Report" version 1 (September 2018 - rescheduled December 2018 - cancelled December 2018)

3.1.7. D5 "Results Report" version 1 (July 2019)

The first version of this document presenting the results achieved during the first 19 months of the project.

3.1.8. D1.1 "System Infrastructure" version 2 (December 2019 -not required-)

3.1.9. D1.2 "Camera simulation" version 2 (December 2019 -not required-)

3.1.10. D1.3 "Evaluation datasets" version 2 (December 2019 -not required-)

3.1.11. D5 "Results Report" version 2 (March 2020)

This document, updating the results of the Project till March 2020.

3.1.12. D1.1 “System Infrastructure” version 2 (December 2020)

This deliverable describes the updated infrastructure available at VPULab for the development of the tasks proposed within the MobiNetVideo project (TEC2017-88169-R. This update focuses on both the acquired hardware and the software infrastructure that allows data processing for large-datasets and simultaneous remote users.

3.1.13. D1.2 “Camera simulation” version 2 (December 2020)

This deliverable updates the description of the work related with the task T.1.2 Cameras network simulation which supports other tasks for generating test data. We focus on the improvements of the simulator “Multi-camera System Simulator (MSS)”: semantic map generation of camera views, new camera types and new Python API.

3.1.14. D1.3 “Evaluation datasets” version 3 (December 2020, not required)

3.1.15. D5 “Results Report” version 3 (December 2020)

This document, updating the results of the Project till December 2020.

3.1.16. D3 “Technologies for mobile camera networks” version 2 (June 2020 – rescheduled December 20202 – rescheduled February 2021)

This deliverable describes the work related with tasks T.3.1 Scene Recognition, T3.2 Semantic Segmentation, T3.3 Multi-view matching and T.3.4 Cooperative detection and tracking

3.1.17. D4 “Deployment and application scenarios” (December 2020 – rescheduled September 2021)

This deliverable recapitulates the different applications developed within WP4 during the 9 months extension of the project (January-September 2021).

3.1.18. D5 “Results Report” version 4 (September 2021)

This document, updates the results of the Project during the extension period (January-September 2021).

3.2. Technical Reports

4. Public Resources

4.1. Content Sets

- **P365LLds**: A Places365 Lifelogging version Dataset. *(available on-demand; to be available on-line soon)*
  The task of scene recognition has been classically evaluated using still images representing scenes. In the context of the MobiNetVideo project, we have created a new dataset that extrapolates Places365's classes to lifelogging/egocentric videos. The dataset is made up of 450 videos recorded with smartphones, go-pro and handheld cameras. Videos have been obtained by downloading YouTube videos licensed as Creative Commons. For each scene class in Places365, we include between one (90% of the classes) and four videos. The average length of the videos is 638 frames (around twenty-one seconds) and the median length is 600 frames per video (around twenty seconds). In overall, the dataset is approximately 34.1 GB large.
- **USSds**: A Unified Semantic Segmentation Dataset *(available on-demand; to be available on-line soon)*
  There is a large variety of semantic datasets. However, not all of them have the same semantic classes, and the appearance of shared classes substantially differ. The USSds represents a data integration effort to create a unified semantic dataset which—by enlarging the number of classes and the diversity of the shared classes, aims to provide a more generic benchmark for training and evaluation. The merged datasets have been relabelled to a common set of 293 semantic labels distributed into a total of 145,555 training images and 7,614 validation images. The datasets agglutinated to compose the USSds dataset are:
  - COCO-Stuff, COCO-Stuff Dataset.
  - Cityscapes, Cityscapes Dataset.
  - ADE20K, ADE20K Dataset.
  - TASKONOMY, TASKONOMY Dataset.
  - Mapillary, Mapillary Dataset.

4.2. Software

- Pytorch Implementation of Semantic-Aware Scene Recognition [1] https://github.com/vpulab/Semantic-Aware-Scene-Recognition.
- Support for Pytorch and Caffe CNNs Visualization [11] https://github.com/MEscuderoVinolo/MobiNet-Video-CNN-Visualization.
- Pytorch/Tensorflow implementation for running multiple detectors and tracking algorithms[26]
  https://github.com/JorgeMunnozAguado/MOT-experimental-framework

5. Workshops and Seminars

5.1. MobiNetVideo Industry Day

The MobiNetVideo Industrial Day was held on June 24th, 2021. Although it was planned as a face-to-face event, finaly it was held on-line in order to accommodate the suggestions of the attending companies.
Different applications and demonstrators developed within WP4 were presented:

- Continual learning for object detection
- Places detection in ego-centric video
- Contributions to the re-identification of objects
- Multi-Object Tracking and Applications
- Real-time camera operation and tracking for the streaming of teaching activities
- Computer Vision Applications for Event Cameras

After the presentatios, there was a discussion session with the industry representatives in order to get feedback, mainly with respect to suggestions for enhancements and new applications

The Industry participation included people from the following companies:

- Thales Aelnia Space
- Nokia Bell Labs Madrid
- Treelogic
- Vaelsys
- Dive Tech
- Capgemini

6. Main Achievements of the Project

6.1. First 18 months main achievements

6.1.1. WP1: Infrastructure and datasets

- T1.1: Infrastructure update and maintenance
  - Acquisition and configuration of new hardware for GPU-based processing of image and video signals (Deliverable D1.1)
  - Acquisition and configuration of new hardware for research activities on "People Tracking for Active Vision" of WP2 and WP3 (Deliverable D1.1)
- T1.2: Cameras network simulation
  - Integration of previous research on simulation of camera networks using the Unity3D game engine. Generation of documentation for the integrated project. (Deliverable D1.2)
  - Design of an API using Python and Matlab. On-going Implementation and testing.
- T1.3: Generation of datasets
  - Development of a tool for the extraction of urban images through predefined routes, using the Google Directions and Street View APIs [27].
  - Compilation of existing datasets for multi-target tracking from UAV-based video sequences. (Deliverable D1.3)
  - Compilation of existing datasets for scene recognition in still images. (Deliverable D1.3)
  - Aggregation of existing datasets for semantic segmentation, creating a unified dataset. (Deliverable D1.3)
  - Creation of a new dataset datasets for scene recognition in life-logging videos. (Deliverable D1.3)

6.1.2. WP.2: Feasibility studies

- T2.1: People tracking for active vision
  - Adaptation of a people detection system for omnidirectional cameras to deep learning descriptors [28].
- T2.2: Object detection for collision detection
  - Evaluation of state-of-the-art in people detection methods with a wearable camera. (Deliverable D2.1).
  - Evaluation of state-of-the-art in people re-identification methods [30].
  - Evaluation of state-of-the-art in car re-identification methods [10].
  - Development of a people detection approach in presence of groups with CNNs [29].
- T2.3: Scene categorization for life-logging
  - Evaluation of state-of-the-art object detectors in urban images obtained with the tool developed in T1.3 [27].

- Evaluation of state-of-the-art scene recognition methods in still images [9].
- Evaluation of state-of-the-art semantic segmentation methods in life-logging videos compiled in T1.3. A Graphical User Interface has been designed to ease results' understanding. (Deliverable D2.3).
- Evaluation of state-of-the-art scene recognition methods in life-logging videos compiled in T1.3. A Graphical User Interface has been designed to ease results' understanding. (Deliverable D2.3).
- Adaptation of a convolutional neural network visualization tool for accounting scene recognition learning process. (Deliverable D2.3).

- T2.4: Multi-target tracking for UAV monitoring
  - Generation of a framework for evaluating algorithms multi-target tracking (developed in Python or Matlab) on existing datasets or custom datasets and using standard evaluation protocols (Deliverable D2.1).
  - Evaluation of selected state-of-the-art approaches using the VisDrone dataset and CarPK dataset captured from drones (Deliverable D2.1).

6.1.3. WP.3: Enabling technologies

- T3.1: Scene identification
  - Development of a framework for scene recognition in Python which combines semantic and appearance deep features by means of an attention module in an end-to-end multi-modal CNN.
  - Implementation of evaluation protocols to estimate scene recognition performance.
- T3.2: Semantic segmentation
  - Development of a framework for semantic segmentation in Python to allow the learning of the unified semantic segmentation dataset.
- T3.3: Multi-view matching
  - Development of a framework for multi-camera object re-identification
  - Implementation of evaluation protocols to estimate object re-identification performance
  - Participation [3][10] in the AICity challenge 2019 (CVPR workshop Challenge Track 2: City-Scale Multi-Camera Vehicle Re-Identification) https://www.aicitychallenge.org/2019-data-sets/.
- T3.4: Cooperative detection and tracking
  - Development of a framework for single-camera Multiple-Object-Tracking in Python which employs spatial and appearance (deep) features.
  - Implementation of evaluation protocols to estimate single camera multiple-object tracking performance.
  - Participation[2] in the VisDrone challenge 2018 (ECCV workshop) http://aiskyeye.com/.
  - Development of a framework for multi-camera Multiple-Object-Tracking in Python which employs spatial and appearance (deep) features.
  - Implementation of evaluation protocols to estimate multicamera multiple-object tracking performance

- Participation [3] in the AICity challenge 2019 (CVPR workshop) https://www.aicitychallenge.org/2019-data-sets/.

6.1.4. WP.4: Deployment of applications

No activity was planned during the first 18 months of the project in this workpackage tasks.

- T4.1: Case studies: implementation and testing
- T4.2: Real-time and network issues
- T4.3 Sensor issues

6.1.5. WP.5: Management and dissemination

- T5.1: Management
  - During the first 18 month, several workplan updates have been done due to the disarrangements caused by the delay in the project kick-off and the resolution of the FPI scholarship.
- T5.2: Dissemination
  - The web page was created and it is being updated timely.
  - Three MobiNetVideo Newsletters were published.
  - As expected, the number of results ready for publication during this period is limited. A detailed plan will be done early September.
- T5.3 Workshop organization
  - The 2018 Developers Workshop was cancelled. Currently, we are evaluating the planning of such workshops for the rest of the project.

6.2. Main achievements during months 19 to 27

6.2.1. WP1: Infrastructure and datasets

- T1.1: Infrastructure update and maintenance
  - Design and Development of a virtualization platform to enable the research and prototyping algorithms with shared computation and storage resources. It also allows concurrent access and remote working by multiple users. On-going documentation and testing.
- T1.2: Cameras network simulation
  - Update of the simulator backbone software (Unity 3D) to latest version.
  - Update of the Python API to Python version 3.7. Completed testing and examples of usage.
  - Addition of new features: cameras of the scene, PTZ cameras, continuous/on-demand running modes and semantic maps for camera feeds. On-going documentation and testing.
- T1.3: Generation of datasets
  - Creation of the **P365LLds** dataset for scene recognition in life-logging videos.
  - Creation of the **USSds** dataset to increase the generality and diversity in the training and evaluation of semantic segmentation methods.

6.2.2. WP.2: Feasibility studies

- T2.1: People tracking for active vision
  - Development of a people detection system for omnidirectional cameras based on an end-to-end deep learning architecture (ongoing work)
  - Adaptation of a state-of-the-art real-time tracker to enhance its robustness. Participation into the Visual Object Tracking 2020 Challenge, specifically into the short-term real-time challenge.
- T2.2: Object detection for collision detection
  - Development of an algorithm for automatic anticipating accidents or collisions in board first-person view camera videos (ongoing work).
- T2.3: Scene categorization for life-logging
  - Statistical analysis of the increase in shared semantic class divergence provided by the **USSds** semantic segmentation dataset (ongoing work).
  - Evaluation of state-of-the-art semantic segmentation urban images obtained with the tool developed in T1.3 [27] (ongoing work).
  - Design, development and limited evaluation of a transfer learning method that moderately conserves previous learning in a CNN trained for scene recognition [11].
- T2.4: Multi-target tracking for UAV monitoring

6.2.3. WP.3: Enabling technologies

- T3.1: Scene identification
  - Design, development and evaluation of a scene-recognition method that takes advantage of the fact that the learning of scenes is inherent to the learning of the objects they include to improve scene-recognition by relying on semantic-segmentation with moderate increasing of the CNN capacity [1].
  - Design and development of computer vision methods for the temporal arrangement and clustering of video sequences based on scene-recognition results (ongoing work).
  - Scene recognition using deep neural networks trained with the Places database [31].
- T3.2: Semantic segmentation
  - Design, development and (limited) evaluation of self-taught and self-paced methods for training severe class-unbalanced datasets as the **USSds** semantic segmentation dataset (ongoing work).
- T3.3: Multi-view matching
  - Work towards participating in the AICity challenge 2020 https://www.aicitychallenge.org/
- T3.4: Cooperative detection and tracking
  - Development of an algorithm for cooperative detection of moving targets in multicamera scenarios based on Deep Features and Mixup learning. On-going documentation and testing.
  - Integration of cooperative detection for target tracking with multiple overlapped cameras. On-going documentation and testing.

6.2.4. WP.4: Deployment of applications

WP4 was planned for starting at month 24, but in order to concentrate resources in WP3, we have decided to delay WP4 kick-off till the planned end of WP3 (M30 – June 2020). Therefore, W4 will start July 2020 (M31).

- T4.1: Case studies: implementation and testing
- T4.2: Real-time and network issues
- T4.3 Sensor issues

6.2.5. WP.5: Management and dissemination

- T5.1: Management
  - During months 19-27 (June 2019-March 2020), there were no major management issues besides the delay in the kick-off of WP4 and the very limited mobility (research stays and conferences).
  - Taking into account the project progress, before summer we will evaluate the possibility of asking for an extension in order to be able to fulfil properly WP4 activities as well as further dissemination of the project.
- T5.2: Dissemination
  - The web page was updated timely.
  - The fourth MobiNetVideo Newsletter was published December 2020.
- T5.3 Workshop organization
  - Developers Workshops are cancelled. Currently, we are evaluating the planning of a final workshop for the disseminating project results to EPOs and industry.

6.3. Main achievements during months 28 to 36

6.3.1. WP1: Infrastructure and datasets

- T1.1: Infrastructure update and maintenance
  - Acquisition and configuration of new hardware for GPU-based processing of image and video signals (Deliverable D1.1 v2)
  - Description of the software infrastructure for multiple remote users and large data processing (Deliverable D1.1 v2)
- T1.2: Cameras network simulation
  - Design and final testing of an API using Python. Examples and documentation generative. (Deliverable D1.2 v2)
  - Protocol and software for automatic generation of segmentation maps for cameras views in the simulator (Deliverable D1.2 v2)
- T1.3: Generation of datasets
  - Nothing to report.

6.3.2. WP.2: Feasibility studies

This Workpackage was finished in the previous reported period.

6.3.3. WP.3: Enabling technologies

- T3.1: Scene identification

  - Advances on scene recognition by combining colour, depth and semantic modalities [15].
  - Image classification with deep neural networks through reduced training sets and few-shot learning [32]
  - Unsupervised scene recognition using features extracted from pre-trained neural networks [33]
- T3.2: Semantic segmentation
  - Advances on the analysis of semantic segmentation methods in the unified dataset. Further tests and documentation (Deliverable D3 D3v2).
  - Development of a framework for semantic segmentation in urban-image databases [38].
- T3.3: Multi-view matching
  - Participation [4] in the AICity challenge 2020 (CVPR workshop Challenge Track 2: City-Scale Multi-Camera Vehicle Re-Identification) https://www.aicitychallenge.org/2020-data-sets/.
  - Study of people and vehicle re-identification systems [37] based on the combination of deep learning characteristics and traditional characteristics that describe the data used, named "visual attributes".
- T3.4: Cooperative detection and tracking
  - Participation [12][6] in the VOT2020 Challenge (ECCV workshop) https://www.votchallenge.net/vot2020/
  - Improvement of the framework for multi-camera Multiple-Object-Tracking in Python which employs spatial and appearance deep features (Deliverable D3 D3v2). Further tests and documentation.
  - Prospective study on action recognition in lifelogging videos [5] and participation in the Sixth International Workshop on Egocentric Perception, Interaction and Computing (EPIC@CVPR2020) workshop.
  - Prospective study on spatio-temporal segmentation based on deep features of life-logging sequences recorded from moving vehicles [36].
  - Development of a people detection system for omnidirectional cameras based on an end-to-end deep learning architecture [13]

6.3.4. WP.4: Deployment of applications

WP4 was rescheduled again, starting finally January 2021. This will be the main work focus during the project extension.

6.3.5. WP.5: Management and dissemination

- T5.1: Management
  - During months 28-36 (April-December 2020), the major management issue has been maintaining activity during these months taking into account the global world situation; several deliverables were delayed. The extension was requested and accepted November 2020, extending the project till September 2021, in order to complete publications and dissemination to industry within WP4. Several adjustments were done after extension approval.
- T5.2: Dissemination
  - The web page was updated timely.

  - The fifth and sixth MobiNetVideo Newsletters were published June and December 2020.
- T5.3 Workshop organization
  - We are working towards a workshop for the disseminating project results to EPOs and industry, as well as obtaining clues and feedback for WP4 activities.

6.4. Main achievements during months 37 to 44

6.4.1. WP1: Infrastructure and datasets

- T1.1: Infrastructure update and maintenance
  - Acquisition of new hardware (disks) for maintenance tasks.
- T1.2: Cameras network simulation
  - Integration of a methodology to evaluate detection algorithms. Reorganization of semantic annotations for available scenarios [47].
- T1.3: Generation of datasets
  - Nothing to report.

6.4.2. WP.2: Feasibility studies

This Workpackage was finished in the previous reported period.

6.4.3. WP.3: Enabling technologies

- T3.1: Scene identification
  - Development of an attribution method to visualize and interpret the predictions of a scene identification model with respect to the input image using semantic information [8].
  - Execution of neural networks on Android smartphones with hardware acceleration through Keras and TensorFlow Lite with application to real time object detection and scene recognition [43].
  - Scene recognition through multi-scale integration of convolutional neural networks [44].
- T3.2: Semantic segmentation
  - Development of a semantic segmentation algorithm focused in a binary objective; differentiate between a specific class and the rest of the image, thus allowing us to differentiate between a specific object and the background [40].
  - Explore the use of semantic segmentation with the objective of improving the performance of deep-learning reinforcement models, and in particular, the performance over a videogame [17].
  - Quantification of the diversity provided by the unified semantic segmentation dataset [22].
  - Evaluation of sample ordering functions to cope with sample unbalance in the training of semantic segmentation models [23].
  - Exploration of the use of semantic segmentation information to establish detection thresholds [42].

  - Exploration of the use of synthetic data for semantic segmentation to enrich the training process [21][42].
  -
- T3.3: Multi-view matching
  - Participation [7][19] in the AICity challenge 2021 (CVPR workshop Challenge Track 2: City-Scale Multi-Camera Vehicle Re-Identification) https://www.aicitychallenge.org/2021-data-and-evaluation/.
  - Prospective study and use of algorithms or continuous learning techniques for different sets of data and images [39][25].
- T3.4: Cooperative detection and tracking
  - Development of an on-board computer vision based accident detector around a moving vehicle [41].
  - Development of a method to cope with occlusions and distractors in tracking methods based on Siamese CNN and correlation filters [20].

6.4.4. WP.4: Deployment of applications

WP4 was rescheduled again, starting finally January 2021. This will be the main work focus during the project extension.

- T4.1: Case studies: implementation and testing
  - Development of vehicle re-identification application (Deliverable D4) able to perform and evaluate the proposed vehicle re-identification system [7][19].
  - Development of a final user interface able to evaluate different continual learning alternatives [25], the user can train and test the model and see the results without using the console (Deliverable D4).
  - Development of a place reidentification application to spatiotemporally locate and associate specific areas of a place in life-logging recordings and of a user interface associate areas in a place recorded in two different moments [18] (Deliverable D4).
  - Development of multiple-object-tracking application (Deliverable D4) able to apply and evaluate algorithms and performance metrics for wearable cameras and moving cameras [26]

6.4.5. WP.5: Management and dissemination

- T5.1: Management
  - During the extension period (January-September 2021), the major management issue has been focused in order to complete publications and dissemination to industry within WP4. Deliverables D3,.v2, D4 and D5.4 (this document) were published.
- T5.2: Dissemination
  - The web page was updated timely.
  - The seventh MobiNetVideo Newsletters was published September 2021.
- T5.3 Workshop organization
  - Industry Day (see section 5.1) was held June 24th