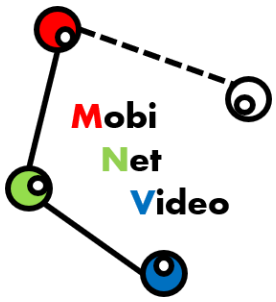

MobiNetVideo Newsletters

#6 – December 2020

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

Visual Analysis for Practical Deployment of Cooperative Mobile Camera Networks

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

Project extension approved

On November 11, 2020, MobiNetVideo extension was approved (till September 30th, 2021).

The main objective during the extension is working in the activities planned with WP4, that is, the development of applications and demos making use of the algorithms developed in the project. These applications will contribute to disseminate the results of the project to companies. During the first months of 2021, a workshop will be held to discuss with Observers and other companies the possible applications of interest.

Additionally, some work will be done within WP3 in order to publish the results of the project in international journals and conferences, as well as publishing more public resources (datasets and opensource).

The web site will be maintained and whilst D4 has been delayed till September 2021, in that date the Final Results Report (D5v4) will be issued, as well as MobiNetVIdeo Newsletter #7 (both new publications due to the extension).

Sixth semester progress report

During this semester, research activities have been progressing properly, still in an almost complete online working environment (remote home working, teleconferencing and mail). Two conference papers have been published and several journal papers are under preparation, a couple almost completed and to be submitted next month.

Deliverables D1.1v2 “System Infrastructure”, D1.2v2 “Camera Simulation” and D5v3 “Results Report” have been published December 2020. In principle, no new version of D1.3 was considered necessary and, in principle, we do not expect if neither during the extension. D3v2 “Technologies for mobile camera networks” has been rescheduled February 2021, and D4 “Deployment and applications scenarios” September 2021.

Sixth semester results

Conferences

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 (CVPRW2020), Seattle, Washington, USA, Jun. 2020, in press

Action recognition is currently one of the top challenging research fields in computer vision. Convolutional Neural Networks (CNNs) have significantly boost edits performance but rely on fixed-size spatio-temporal windows of analysis, reducing CNNs temporal receptive fields. Among action recognition datasets, egocentric recorded sequences have become of important relevance while entailing an additional challenge: ego-motion is unavoidably transferred to these sequences. The proposed method aims to cope with it by estimating this ego-motion or camera motion. The estimation is used to temporally partition video sequences into motion-compensated temporal chunks showing the action under stable backgrounds and allowing for a content-driven temporal sampling. A CNN trained in an end-to-end fashion is used to extract temporal features from each chunk, which are late fused. This process leads to the extraction of features from the whole temporal range of an action, increasing the temporal receptive field of the network.

M. Kristan et al., **The Eighth Visual Object Tracking VOT2020 Challenge Results**, ECCV Workshops 2020, Online, Aug. 2020, in press

The Visual Object Tracking challenge VOT2020 is the eighth annual tracker benchmarking activity organized by the VOT initiative. Results of 58 trackers are presented; many are state-of-the-art trackers published at major computer vision conferences or in journals in the recent years. The VOT2020 challenge was composed of five sub-challenges focusing on different tracking domains: (i) VOT-ST2020 challenge focused on short-term tracking in RGB, (ii) VOT-RT2020 challenge focused on “real-time" short-term tracking in RGB, (iii) VOT-LT2020 focused on long-term tracking namely coping with target disappearance and

reappearance, (iv) VOT-RGBT2020 challenge focused on short-term tracking in RGB and thermal imagery and (v) VOT-RGBD2020 challenge focused on long-term tracking in RGB and depth imagery. Only the VOT-ST2020 datasets were refreshed. A significant novelty is introduction of a new VOT short-term tracking evaluation methodology, and introduction of segmentation ground truth in the VOT-ST2020 challenge - bounding boxes will no longer be used in the VOT-ST challenges. A new VOT Python toolkit that implements all these novelites was introduced. Performance of the tested trackers typically by far exceeds standard baselines. The source code for most of the trackers is publicly available from the VOT page. The dataset, the evaluation kit and the results are publicly available at the challenge website.

Master thesis

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.

Abstract: The implementation of a system oriented to the study of artificial vision is interesting due to the relevance it can have on the development of simulations, which allows the creation of scenes and events that contribute to the continuous improvement of code and algorithms within artificial vision. Based on the API (Application Programming Interface) of the simulator called: Multi-camera System Simulator (MSS) developed in the TFG of 2017 by Mario González within the VPULab research laboratory of the Universidad Autónoma de Madrid, the present End of Master's Work (TFM) incorporates improvements in the simulator to enhance its functionality. The project provides library updates, new functions and a client API developed in Python (v3.6.9). This work makes use and handling of multiple virtual cameras implemented in the first part by the scenario designer and on the second part through a user API that incorporates a wide range of functionalities for handling the system with multiple cameras. The tool allows the configuration of the cameras and their use both remotely and locally with a client-server structure. In addition, the document details the design, implementation, integration and changes required to generate unlimited test environments. Finally, the performance of the developed system is evaluated and the results are discussed, establishing recommended configurations and discovering the technical limits of the simulator, as well as an evaluation of the behaviour of the computer resources when the simulator is executed.

Graduate thesis

Reconocimiento no-supervisado de escenas mediante características extraídas de redes neuronales pre-entrenadas (**Unsupervised scenes 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.

Abstract: Neuronal networks have been growing enormously from the moment that could be considered as the beginning of this technology with the "perceptron" and the "multilayer perceptron" ïnvented in the late 50s and the beginning of the 60s. From that moment on, improvements kept coming, and new approaches of the "backpropagation algorithm" came out. This was an extraordinary revolution that made available returning back to previous layers and obtain the cost function and make small changes that lead to our network to classify correctly the images. In the further years, improvements continued and with them appeared the convolutional neural network, based on the visual cortex of animals. The first papers related to hand-written figures recognition, the "hello World" of this field. In the current century, the technology of GPUs was used and the problems related to computational maths disappeared with them what leads to an almost unlimited power of this networks. All this discoveries and improvements have made this technology to be very present in our daily lives and can me mentioned in fields like medicine, financial markets, automotion and tons of other examples. This paper will focus in place recognition based in feature extraction from a pre-trained neural network. For this, Andreas Sebastian Wolters' paper "Unsupervised scene and place recognition based on features extracted from pre-trained convolutional neural networks" has been followed, where CNNs as AlexNet or VGG16 are used firstly with a matching algorithm for the place recognition where a look-up table formed by different and new places is created and then a k-means clustering algorithm for the scene recognition based on the dictionary generated before. For this project, pre-trained neural networks have been used with the ImageNet dataset that enabled a quick train and the possibility of not using a huge dataset of own images for training our model. Finally, a comparison and analysis have been done taking into account the different measurements used and the results provided.

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.

Abstract: This Final Bachelor Project will deal with deep learning algorithms based on semantic segmentation. The purpose of this project will be to test various semantic segmentation algorithms with video simulations in lieu of real videos. To execute this task, first of all there will be an introduction displaying the technological advances that the eld of image treatment has experienced in the past years, followed by a brief explanation of semantic segmentation. Then, the project will proceed to explain the mechanisms of sematic segmentation before elaborating on the theoretical concepts of the convolutional neural networks. Once the idea of semantic segmentation has been fully drawn out, the project will proceed to individually study each assignment specifically looking at the used architecture as well as the required structural framework. To be able to implement each task of the sequence generation of video simulations, there will be a display of different graphic motors available nowadays and will undeniably justify the use of the graphic motor Unity. These videos will vary in type and will simulate various scenarios including crossroads with both cars and pedestrians simultaneously circulating as well as an aerial view from an helicopter. To materialize each task there will be multiple pre-trained models implemented provided by various assignments, which will individually check their robustness. Once all of the required tests have been made, a performance evaluation of all the implemented algorithms will follow. This will be accompanied by a discussion focusing on each of the obtained results allowing to determine which of them adapts itself the best to the video sequence simulations.

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.

Abstract: This Bachelor Thesis aims at the creation of a complete infrastructure based on several virtualization and distributed system techniques. As a result, a robust and efficient system is achieved, including a modular architecture that simplifies scalability, disaster recovery and data backup. This work develops an integral solution, because of that, all necessary steps until the final implementation are described. Such process entails evaluation of the available alternatives, logic design of systems and network architecture, system installation and required procedures for the implementation of all components. On the other side, a technical documentation has been elaborated for the staff to be capable to adjust the system to their needs by modifying and extending it. In the same way, multiple guides were written which help final users to utilize the services offered in the laboratory so they could quickly be able to get up and

running. This work's results will be used in a daily basis by the people in the research group so it is crucial that the system works as expected, simplifying the researcher's work. For that reason, all components will be evaluated to ensure optimal performance and the absence of anomalies during the process.

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.

Abstract: In this final degree project, the objective is to study the feasibility of spatiotemporal grouping the frames in a video captured by a moving vehicle or a person walking using the visual characteristics extracted by means of a Convolutional Neural Network. The goal is to divide the path followed by the vehicle into groups, paying attention to its changes in orientation or position relying exclusively on visual data captured by an on-board camera. To this aim, a database is designed using images from the Google Street View repository. This database is made up of videos captured by the vehicle while recording trajectories in different places around the world. The so-obtained database is cleaned by removing the images that are very similar to their predecessor, which result in training data with less redundant information. Then, we rely on a GoogleNet pre-trained for image classification. To adapt this network to our database, tests are performed, sweeping different hyperparameter to establish the best network configuration. Once these are found, several training sessions are carried out to automatically classify the videos, varying the learning permeability of their layers, in order to use part of the characteristics learned in previous training sessions. Once the networks are trained, we propose to extract characteristics for each frame of the video, by truncation of the network at different levels. With these characteristics, we calculate the distances between all the frames of each video, obtaining distance matrices. Finally, using distance matrices, multiple tests are performed using different grouping algorithms, distance evaluation metrics and criteria for clustering frames. In these tests the goodness of the clusters obtained for each of the trajectories is evaluated through the execution of prospective cases. Preliminary results suggest that the designed method can obtain spatiotemporal groupings of the video frames that approximately agree with the human interpretation of the trajectory followed by the capturing camera.

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.

Abstract: The main focus of this final degree project is the study of people and vehicle re-identification systems based on the combination of deep learning characteristics and traditional characteristics that describe the data used. First, the state of the art will be studied to understand the deep learning techniques based on neural networks that will be essential to extract the necessary characteristics in re-identification tasks. Secondly, the data sets used and their respective attributes for both people and vehicles will be described. In addition, a short introduction to the development environment used (Pytorch) and the pre-trained models chosen to carry out the experiments will be included. Once the previous concepts have been explained, a summary of the implemented design will be made from them. Finally, the metrics used to evaluate and extract re-identification results will be explained, as well as different techniques to improve said results. Numerous experiments will be carried out using several trained models with different parameters to draw general conclusions from the results and then perform the combination of the characteristics to observe the effect on the final result. Before finishing the final project, the results obtained will be evaluated and visual tests will be provided to demonstrate the effect achieved, and to finish, new research routes will be proposed that could be carried out in the future.

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