MobiNetVideo Newsletters

#3 - July 2019

TEC2017-88169-R MobiNetVideo (2018-2020)

Visual Analysis for Practical Deployment of Cooperative Mobile Camera Networks

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

Third semester progress report

During this third semester, the first set of deliverables (D1.1 “System Infrastructure” version 1, D1.2 “Camera simulation” version 1, D1.3 “Evaluation datasets” version 1, D2 “Feasibility studies: algorithms and findings”, D5 “Results Report” version 1) as well as Newsletter#3 were finally rescheduled and published July 2019. Deliverable D3 “Technologies for mobile camera networks” version 1, has been rescheduled for September 2019.

Research activities have been progressing properly. Considering the Results Report (D5v1), we will work on potential new changes to the workplan and project schedule during September.

Third semester results

Conferences

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 (CVPR2019), Long Beach, CA, USA, Jun. 2019 (in press)

Abstract: In this paper, we present an approach for Multi-target and Multi-Camera Vehicle Tracking and another approach for Vehicle Re-Identification (ReID) across multiple cameras. We evaluate both approaches over ”CityFlow: A City-Scale Benchmark” participating in Track 1 and Track 2 of 2019 AI City Challenge Workshop. The proposed tracking approach is based on applying detection and tracking of multiple moving vehicles for each camera. Afterwards, we cluster such results (detections of vehicles) obtained from multiple cameras with overlapped fields of view. The clustering is based on appearance and spatial distances on a common plane for all camera views. The optimal number of clusters is obtained by using validation indexes. Then, a spatio-temporal linkage of the obtained clusters is performed to obtain the trajectories of each moving vehicle in the scene. We tested different combinations for the input of the proposed approach (detector and tracker) and provide sample results for selected scenarios of ”CityFlow: A City-Scale Benchmark”. The proposed re-identification system is based on the combination of adapted deep learning feature embedding representations and a distance metric learning process. We also include the vehicle tracking information provided by the ”CityFlow: A City-Scale Benchmark” in order to improve the results. We tested different combinations of features, metric learning and the use of tracking information and provide sample results for the CityFlow-ReID dataset.

Master thesis

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.

Abstract: The detection and association of objects in multiview scenarios is an area of research within the Computer Vision that is very useful in tasks such as video surveillance, for example when identifying in different scenes a person who has carried out any kind of anomaly. This project is going to focus on vehicle re-identification, due to it has been a critical problem in the Intelligent Transportation System (ITS) for the recent years, and we can see our performance compared with the state of the art participating in the 2019 AI City Challenge. The main objective of this Master Thesis is the development of a system that detects and associates multiple objects in multiview scenarios based on deep learning. For this task, different algorithms from the state of the art have been studied. Furthermore, the method implemented uses feature extraction methods and metric learning techniques and it returns a list with all the matches between the query object and the images from the gallery set, sorted according to their distance. A new dataset has been reorganized from the train part of the CityFlow-ReID dataset. In order to improve the results, it is included a metric network combination at distances and ranks level from the different feature extraction methods and metric learning techniques. Finally, we have developed our own experimental setup.

Graduate thesis

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), Trabaja Fin de Grado (Graduate Thesis), Grado en Ingeniería de Tecnologías y Servicios de Telecomunicación, Univ. Autónoma de Madrid, Jul. 2019.

Abstract: In this Graduate Thesis proposes two objectives: the development of an algorithm for the creation of a data base formed through the extraction of images from Google Street View along different routes and testing and evaluating an object detector over these images. In the first place, a tool which extract images along selected routes is developed. These routes are the optimal ones, which are calculated trough the Google Maps algorithms, from an origin, a destiny and the transport mode to be used. The extracted images are not panoramic ones, but they allow a variation of parameters. Some of these parameters are the angle from which the images are taken, horizontal (heading) as well as vertical (pitch), and the field of view (FoV), to be able to represent the whole surroundings where the images are taken. In this way, the images with different objects, characteristics and distortions are collected. In the second place, once the data base with our images is created, the object detection model of YOLOv3, pretrained with the data base COCO, is applied to it, in order to detect certain objects in them. From these results an algorithm is created to the optimal representation of these detections on the respective images. Finally, in order to complete the objective, we implemented a system to evaluate the quality of the detection on the images, through the calculation and representation of the relation between the precision and the recall of the model. To this aim, we create a Ground Truth over some images as well as a program which compares it with the detections produced by the detector over these images. Besides, some cases in which the detector was not efficient enough, because of some images have distorted areas and objects, are studied and evaluated.

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), Trabaja Fin de Grado (Graduate Thesis), Grado en Ingeniería de Tecnologías y Servicios de Telecomunicación, Univ. Autónoma de Madrid, Jul. 2019.

Abstract: In the field of computer vision, convolutional neural networks (CNNs) are highly effective in object detection and classification, due to their ability to process information from almost any type of data that can be represented in a two-dimensional space, such as images or even music. This Project tries to improve the efficiency of a real-time people detection system with omnidirectional cameras and its generalization ability, thanks to the advantages of pre-trained convolutional neural networks as compared to classic image descriptors. Due to the fact that training a network is a slow and expensive process, we use a pre-trained network which offers satisfactory results and eliminates the requirement to adjust or re-train the network. Therefore, the objective is to include in the system the capability to extract descriptors from pretrained CNN. The classifier of the people detection system consists of a support vector machines (SVM) grid distributed throughout the image, with an area of action called fovea. As a result, the classifier is able to distinguish the different types of distortion that take place in an omnidirectional image according to the spatial location of an object or, in such a case, a person. The final results have been achieved by extracting the descriptors from some of the layers using some of the most popular CNN pre-trained neural networks, as well as AlexNet, Densenet201, MobileNetV2 and VGG19. These neural networks have been trained with databases with millions of images and thousands of classes, such as ImageNet, making it possible to use them in this type of applications. Finally, the results of each of the networks and layers will be analyzed and the performance of the system will be measured with these descriptors.

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

Abstract: At present, it is not uncommon to feel watched by cameras while we are making normal life by the streets or in closed places. A camera that records 24 hours a day, 365 days a year, generates information of which, if we know how to treat it, we obtain a large amount of valuable data. Nowadays, there are some

people who have worried about the treatment of this data, generating algorithms that allow detect the people who pass in front of the cameras, although that algorithms are already sophisticated, all of them have the same problem, the occlusion between persons at the captured images. Thus, they have been developed algorithms that improves the capture of this people, even though the results are satisfactory, there aren´t enough. Also, in the last years, the use of the Artificial Intelligence, in concrete, Convolutional Neuronal Networks (CNNs) for the treatment of images, has managed to improve these algorithms, which usually used traditional techniques such as gradient descent. The results are very encouraging in all areas of detection such as objects, people, and other elements in images. This is why we propose in this Graduate Thesis to improve an algorithm based on Deformable Parts Model (DPM) and CNNs, with the aim of adding those cases in which we find occlusion of people by others, using a model that not only focuses on the main person but also analyzes its closest environment in search of more persons that may be occluded and in which we only see, for example, the right half of the body. The main objective is to improve the existing algorithm, evaluated through certain representative video sequences being the result satisfactory.

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

Abstract: Nowadays, person re-identification is a resource with a high demand especially in the field of video surveillance, this does not mean knowing the identity of a person but being able to monitor it in different cameras whose images are not overlapped. There are a lot of traditional evaluation measures that allow us to make "manual" feature extractions, which use mathematical algorithms which allow extracting information from the images. However, there are many aspects to consider if we want our models to work as well as possible, such as the orientation of the person, the environment or their position. Feature extraction based on deep learning makes a modeling of data, for this task, uses large data flows to learn from these and to perform a classification and predictive analysis. Deep learning is based on the use of neural networks, a mathematical model that tries to imitate the biological behavior of the neurons, connecting different layers of processing and granting different weights to each in order to obtain an optimized model. The task of this Graduate Thesis is the comparison of the results of the manual extraction measures (handcrafted features) and the automatic ones (based on Deep Learning), this will be done by executing a script that calculates the percentage of success when it comes to identify people

between the cameras using the different methods of manual extraction and then using those based on neural networks in the datasets we use to evaluate.

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

GOBIERNO DE ESPAÑA
MINISTERIO DE ECONOMÍA, INDUSTRIA Y COMPETITIVIDAD