Abstract

The field of computer vision has had great progress during the last decades due to the advancement of hardware computing power, specifically in Graphics Processing Unit (GPU). Although GPSs have been design for the gaming industry, also have been useful to design powerful algorithms for solving some problems such in this field research such as segmentation, detection, structure from motion, camera calibration and many others. These problems are common in applications like autonomous driving, robot navigation, video surveillance for human tracking, action recognition or human body pose estimation. Some techniques has been applied to tackle these tasks, one of the most widely used deep learning based technique, is the convolutional neural network (CNN) due to the power for features extraction in images.

This dissertation presents a series of CNN-based techniques applied to images to tackle the camera pose and human body pose estimation problems from multi-view environments. For the camera pose estimation, two approaches based on Siamese architecture have been proposed to estimate the camera pose—extrinsic parameters. The first approach takes as input a set of pairs of real-images, which should have a minimum overlapping to ensure that the pairs of images have common features. However, due to few real-image datasets available to the camera pose estimation from multi-view scenarios, a second approach is proposed. It consists of domain adaptation strategy, including the generation of different virtual scenarios by using a special 3D simulation software. The strategy is used to take advantage of transferring of learned knowledge from these virtual scenarios to real-world scenarios. For the human body pose estimation problem, two approaches are also proposed. The first, an architecture based on convolutional neural network, which leverages the estimated extrinsic parameters to establish the relationship between different cameras into the multi-view scheme. It has allowed to estimate the human body pose using information from different points of view, and thus, to solve the challenging problem of the self-occlusion in human pose estimation due to the natural body pose. A second approach for the human body pose estimation problem has been also proposed. It uses attention modules to detect body joints. However, unlike to the first approach, this new approach does not take into account the extrinsic parameters between different cameras of the multi-view scheme, instead, the position and orientation of bones of human body are used as additional information to tackle the problem of self-occlusion of human body joints. The accuracy of these estimations is important to avoid possible false alarms in behavioral analysis systems of smart cities as well as applications for physical therapy, safe moving assistance for the elderly among other.