Abstract Human Pose Estimation (HPE) is a challenging area of research in the field of collaborative human-robot systems, where safety remains one of the most critical aspects. The goal is to distinguish and analyze the human pose based on anatomical features extracted from wearable sensors or less intrusive visual sensing modalities. The limitation of conventional HAR methods based on visual sensors is that they adopt RGB images as input to estimate anatomical keypoints with Deep Convolutional Neural Networks (DNNs). However, those approaches neglect the challenge of detecting features reliably during night-time or in difficult lighting conditions, leading to safety issues. In response to this challenge, we compare some of the most popular detection networks and we study how they perform in the task of extracting features under variability of lighting conditions. Given the results of our study, we propose a novel RGB/Infra-Red camera fusion approach, based on the open-source library OpenPose, and we show how the fusion of keypoints extracted from different images can be used to improve the human pose estimation performance in sparse light environments. Specifically, OpenPose is used to extract body joints from RGB and Infra-Red images and the contribution of each frame is combined by a fusion step. We investigate the potential of a fusion framework based on Deep Neural Networks and we compare it to a linear weighted average method. The proposed approach shows promising performances, with the best result outperforming conventional methods by a factor 1.8x on a custom data set of Infra-Red and RGB images captured in poor light conditions, where it is hard to recognize people even by human inspection.

Viviana Crescitelli, Hitachi Ltd., Research and Development Group