Robotics and Autonomous Systems ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Robotics and Autonomous Systems

journal homepage: www.elsevier.com/locate/robot

OpenPTrack: Open source multi-camera calibration and people tracking for RGB-D camera networks

Matteo Munaro *, Filippo Basso, Emanuele Menegatti

Department of Information Engineering, University of Padova, Via Gradenigo 6B, 35131 Padova, Italy

HIGHLIGHTS

- Real-time, easy to use, and precise multi-camera calibration process.
- Calibration refinement method that exploits people detections from every sensor.
- Quantitative comparison of three RGB-D sensors for the purpose of people tracking.
- People detection invariant to light conditions and low-weight tracking algorithm.
- All the software described is publicly available at http://openptrack.org.

ARTICLE INFO

Article history:
Available online xxxx

Keywords:
OpenPTrack
People tracking
RGB-D
Open source
Multi-camera
Network calibration
Human–robot interaction
Microsoft Kinect
Mesa SwissRanger
Stereo

ABSTRACT

OpenPTrack is an open source software for multi-camera calibration and people tracking in RGB-D camera networks. It allows to track people in big volumes at sensor frame rate and currently supports a heterogeneous set of 3D sensors.

In this work, we describe its user-friendly calibration procedure, which consists of simple steps with real-time feedback that allow to obtain accurate results in estimating the camera poses that are then used for tracking people. On top of a calibration based on moving a checkerboard within the tracking space and on a global optimization of cameras and checkerboards poses, a novel procedure which aligns people detections coming from all sensors in a *x-y-time* space is used for refining camera poses.

While people detection is executed locally, in the machines connected to each sensor, tracking is performed by a single node which takes into account detections from all over the network. Here we detail how a cascade of algorithms working on depth point clouds and color, infrared and disparity images is used to perform people detection from different types of sensors and in any indoor light condition.

We present experiments showing that a considerable improvement can be obtained with the proposed calibration refinement procedure that exploits people detections and we compare Kinect v1, Kinect v2 and Mesa SR4500 performance for people tracking applications. OpenPTrack is based on the *Robot Operating System* and the *Point Cloud Library* and has already been adopted in networks composed of up to ten imagers for interactive arts, education, culture and human–robot interaction applications.

1. Introduction

The ability to detect and track people in real time is useful for a variety of applications: from video surveillance to robotics, from education to art. With the advent of commercially available consumer RGB-D cameras, and continued efforts in computer vision research to improve multi-modal image and point cloud processing, robust person tracking from a single camera with the stability and responsiveness necessary to drive interactive applications is now possible at low cost [1]. However, these sensors usually have a limited field of view and working range, that only allows to track people in a narrow area. Moreover, people tracking is still more prone to errors when a person gets occluded by other people or objects. Exploiting multiple cameras in a network can solve both these issues, but the extension of people tracking algorithms to multi-camera scenarios is not straightforward.

* Corresponding author.
E-mail addresses: matteo.munaro@dei.unipd.it (M. Munaro), filippo.basso@dei.unipd.it (F. Basso), emg@dei.unipd.it (E. Menegatti).

http://dx.doi.org/10.1016/j.robot.2015.10.004
0921-8890/

For this purpose, we created OpenPTrack,[1] an open source, BSD licensed, project launched in 2013 to create a scalable, multi-camera solution for person tracking that specifically aims to support applications in education, art, and culture. This project aims at enabling artists and creators to work with robust real-time person tracking in real-world projects. OpenPTrack aims to support *creative coders* in the arts, culture, and educational sectors who wish to experiment with real-time person tracking as an input for their applications. In order to allow application developers to easily use it for their work, we also focused on providing user-friendly calibration and tracking processes by automating most of the operations that were required to the user. The system allows to use a network of imagers to track the moving centroids (center of mass) of people within a defined area. These data are also provided as a simple JSON-formatted stream via UDP, which can be incorporated into creative coding tools like Max/MSP, Touchdesigner, and Processing, as well as a variety of other software languages and environments.

In addition to the applications listed above, OpenPTrack is also perfectly suited to be adopted in environments where a robot has to interact with a human, such as in a *RoboCup@Home* scenario [2] or when a human and a robot share the same workspace in a factory [3]. The constant and real time monitoring of people positions is a necessary condition to avoid any injury that an industrial robot could cause to a human.

OpenPTrack contains numerous state-of-the-art algorithms and pipelines for RGB and/or depth-based people detection and tracking, and has been created on top of a modular node based architecture, to support the maximum re-use of code and a distributed implementation. In particular, it is built on other open source libraries: the Point Cloud Library (PCL [4]), Open Source Computer Vision (OpenCV [5]) and Robot Operating System (ROS [6]). It currently works with the first and second generation Microsoft Kinect, Mesa Imaging Swissranger SR4500 and custom stereo cameras built with a pair of PointGrey cameras.

The main contributions of this work are:

- a real-time, easy to use, and precise calibration process that also exploits people detections for refining camera poses
- a quantitative comparison of three RGB-D sensors for the purpose of people tracking
- a number of new features and optimizations implemented in OpenPTrack with respect to the people detection and tracking algorithms in [1].

The remainder of the paper is organized as follows: In Section 2, we review the most recent works on RGB-D people detection and tracking and camera networks, while in Section 3 we present the sensors that can be used to build an OpenPTrack network. Section 4 describes the three stages of the procedure implemented in OpenPTrack for calibrating a network of RGB-D sensors in real time. Section 5 gives an overview of our detection algorithms and of the differences with [1], while Section 6 details an efficient version of the tracking node and Section 7 explains our efforts towards system usability. Experiments on network calibration and people tracking with different sensors are reported in Section 8, while some interactive applications exploiting OpenPTrack are described in Section 9. Conclusions are drawn in Section 10.

2. Related work

2.1. People tracking from RGB-D data

Since the introduction of low-cost RGB-D sensors, a number of people detection and tracking algorithms which exploit combined color and depth information have been proposed. Most of these approaches apply a sliding window technique to RGB and/or depth images for people detection [7,8], thus requiring high parallelization with GPUs to obtain real time performance. Moreover, the preferred tracking algorithm in this context is a multi-hypothesis tracking, that is able to recover from failures, but is also computationally expensive, thus hardly usable in real time scenarios with many people.

Recent works [9,10,1,3,11] allow to avoid the sliding window approach that usually leads to analyze thousands of detection windows per image. These new methods exploit the assumption, also adopted by OpenPTrack, that people stand/walk on a ground plane and cluster algorithms on depth data to find a small number of Regions Of Interest (ROIs) which are candidates to contain people and are then classified with more robust and computational demanding algorithms.

An open source implementation of the people detection algorithms in [10,1] is available in the Point Cloud Library [4]. In ROS-Industrial *Human Tracker* repository,[2] these algorithms are combined with other people detectors in a people detection cascade which obtains even more computational efficiency and which is combined with the tracking algorithm described in [3]. Also the code for the GPU-based and the CPU-based versions of the people detection and tracking software in [11] has been recently released.[3]

2.2. Multi-camera people tracking

All of the software described above is targeted to track people from a single RGB-D camera which can move onboard a wheeled robot. However, none of these provides the possibility to perform people tracking in a distributed fashion by exploiting multiple cameras.

Some works exist on people tracking and re-identification from multiple 2D cameras. Some of them assume that the network has been calibrated [12,13], while others perform tracking with uncalibrated cameras [14]. One of the main methods for determining spatial positions of people is to geometrically transform images based on a predetermined ground plane homography [15–17]. Tracking can then be done based on the estimated ground plane positions. Unfortunately, none of these works address the problem of network calibration and they do not provide an open source implementation of the presented methods.

2.3. Multi-camera calibration

The problem of network calibration is not necessarily connected to the people tracking application and concerns a whole branch of works in literature. Auvinet et al. [18] proposed a new method for calibrating multiple depth cameras for body reconstruction by using only depth information. Their algorithm is based on plane intersections and the NTP protocol for data synchronization. The calibration achieves good results: even if the depth error of the sensor is 10 mm, the reconstruction error with three depth cameras is, in the best case, less than 6 mm. A drawback of their implementation is that they have to manually select the plane corners and, above all, they only deal with depth sensors, thus avoiding the possibility to add the color information to the fused data. Another approach to solve the calibration problem is the one proposed by Le and Ng [19]: they jointly calibrate groups of sensors. More specifically, each group is composed by a set of

[1] Website: http://openptrack.org, repository: https://github.com/OpenPTrack/open_ptrack.

[2] https://github.com/ros-industrial/human_tracker/tree/develop.

[3] http://www.vision.rwth-aachen.de/software/realpdt/realpdt.

Please cite this article in press as: M. Munaro, et al., OpenPTrack: Open source multi-camera calibration and people tracking for RGB-D camera networks, Robotics and Autonomous Systems (2015), http://dx.doi.org/10.1016/j.robot.2015.10.004