



HAL
open science

Improving Robot Reactivity to Passers-by with a Faster People Detector

Alhayat Ali Mekonnen, Frédéric Lerasle, Ariane Herbulot

► **To cite this version:**

Alhayat Ali Mekonnen, Frédéric Lerasle, Ariane Herbulot. Improving Robot Reactivity to Passers-by with a Faster People Detector. 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI 2016), ACM; IEEE, Mar 2016, Christchurch, New Zealand. pp.477-478. hal-01296641

HAL Id: hal-01296641

<https://hal.science/hal-01296641v1>

Submitted on 1 Apr 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Improving Robot Reactivity to Passers-by with a Faster People Detector

Alhayat Ali Mekonnen
CNRS, LAAS, 7 avenue du Colonel Roche
F-31400 Toulouse, France
Email: aamekonn@laas.fr

Frédéric Lerasle and Ariane Herbulot
CNRS, LAAS, 7 avenue du Colonel Roche, F-31400 Toulouse, France
Univ de Toulouse, UPS, LAAS, F-31400 Toulouse, France
Email: lerasle@laas.fr, aherbulo@laas.fr

Abstract—Perceptual systems deployed in embedded intelligent systems are handicapped with limited computation resources. As a result, they require explicit computation time consideration during design time. In this paper, we investigate the impact a visual person detector explicitly optimized to minimize computational time has on an exemplar robotic application – navigation in human shared environments. The experimental results attest the improved perceptual frame rate leads to significant improvements in vision system performance and robotic mission success rate.

I. INTRODUCTION

In the literature, various researchers continuously propose ways to improve computer vision algorithm performances – both accuracy and speed. But, it is rare to find works that demonstrate the impact/influence a speed improved algorithm has on a complete system functioning through quantitative evaluation. For example, let us consider the case of a visual person detector which has many applications in human-robot interaction. Plethora of research works are available that showcase many frame rate improving designs and implementations [1]. Unfortunately, to the best of our knowledge, it is difficult to find a work that quantitatively demonstrates the improvements brought on a complete functioning system – e.g. dynamic obstacle avoidance in robotic and automotive applications where passers-by are the obstacles.

In this work, we propose to deploy an optimized visual person detector, by replacing the base line HOG detector [2], in a multi-person tracking-by-detection framework on a demonstrative robotic application and evaluate its impact on a robot’s reactivity (dynamic obstacle avoidance) during navigation in human co-occupied environments.

II. SYSTEM OVERVIEW

A. Passers-by Detection and Tracking

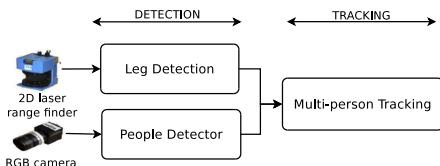


Fig. 1: Multi-sensor people detection and tracking framework

Passers-by detection and tracking in the vicinity of the robot is treated as a multi-person detection and tracking problem. Fig. 1 depicts the corresponding block diagram. It is composed of an RGB camera mounted on a wall (for wider field of

view) and a laser range finder mounted on a mobile robot. People are detected, in both sensors, using dedicated detectors. From the laser data, they are detected with the help of a laser based leg detector that segments laser blobs that conform to the geometrical properties of a human leg.

On the other hand, the detector on the RGB camera is a trained appearance based visual detector. We consider two detector variants, the classical Dalal and Triggs HOG based detector [2] and an optimized variant presented in [3] called BIPBoost. The BIPBoost uses discrete optimization, specifically Binary Integer Programming (BIP), to select a subset of HOG features in a cascade framework that lead to improved frame rate while maintaining comparable detection to the classical HOG detector. Our objective is to assess the effect of these detector choices (HOG vs BIPBoost) on the overall robotic mission performance – specifically, navigation in crowds in an acceptable manner. In our experiments, GPU implementation of these detector variants are used. All the detections are transformed onto a global world frame, providing location on the ground plane and appearance information (based on intrinsic/extrinsic camera calibration, and robot localization module), for proper fusion.

For multi-person tracking, we rely on Reversible Jump Markov Chain Monte Carlo Particle Filter (RJMCMC-PF) [4]. This is a more appealing choice in terms of performance and computational requirement. Our RJMCMC-PF tracker is driven by the detectors, hence a tracking-by-detection approach, that provide ground position of individual persons and their corresponding appearance information. It determines the current position and heading direction of all targets on the ground plane (used to define a personal space for each target).

B. Robotic Integration and Navigation

Many researchers agree on an asymmetric personal space which provides more room up front as people in general are more strict regarding their frontal space [5]. Inspired by this we use the

strict personal space made up of two half ellipses as shown in Fig. 2. For robot navigation, we utilize the Nearness Diagram (ND) navigation [6], which is based on identification of free space areas and obstacles proximity based on some diagrams to define a set of situation which trigger specific motion

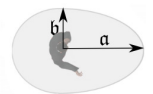


Fig. 2: Personal space model.

laws. In our case, the only modification is: instead of seeing persons as just static obstacles, the algorithm will treat them as obstacles with special zone needs (Fig. 2) that depend on their motion direction – which underlines the need for the tracker.

III. EXPERIMENTS AND RESULTS

All evaluation experiments are carried out on-line using the presented perceptual system. Importantly, two variants are investigated using: (1) classical GPU based HOG detector (GPU-HOG, ≈ 7 fps), and (2) GPU based BIPBoost detector (GPU-BIPBoost, ≈ 12 fps). All tracked targets in the scene wore helmets fitted with reflective markers which are tracked by a motion capture system that provides high accuracy and keeps identity of the marked targets consistently. This is used as a ground truth for evaluations. All in all, we have carried out two types of on-line evaluations.

(1) *Multi-person Tracker evaluation*: where the multi-person tracker is evaluated via live runs using ground truth acquired from a motion capture system. For this, results are reported as Multiple Object Tracking Accuracy (MOTA), and Multiple Object Tracking Precision (MOTP) – the standard tracking evaluation metrics [7] – averaged over several runs comprising a total of more than 2600 frames for each tracker using the different detectors. Three targets are used at most for the evaluation. In general, the perceptual modalities achieve a tracking precision of a little over 20 cms with an accuracy of 72.0% when using GPU-HOG and 74.0% when using GPU-HOG-BIPBoost detector. Under each evaluation criteria, the multi-person tracker based on the GPU-HOG-BIPBoost excels.

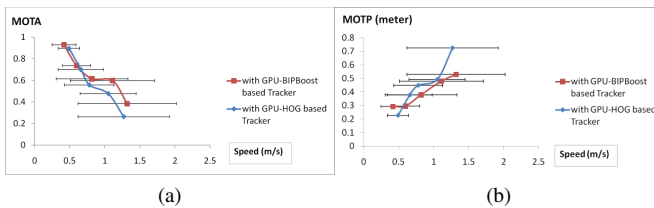


Fig. 3: Evaluation with variable target walking speeds.

To see the effect people’s walking speed has on the tracker (which rely on detectors with different frame rates), we asked the participants to walk at different speeds during each experiment run. Then, multiple runs are evaluated using the GPU-HOG and GPU-HOG-BIPBoost detector variants alternatively. Fig. 3a shows the resulting MOTA as a function of the targets speed and Fig. 3b that of MOTP. For targets’ speed, what we have done is, to measure their speed (obtained using the motion capture system) at each frame and then determine the average speed and its standard deviation using the data of only a complete run. We make sure each run is more than 200 frames to gather characteristic evaluation. In the figures 3a and 3b, a marked point corresponds to a single trial with its x-value corresponding to the average speed and y-value to the measured quantity. The standard deviation shown (in the form of error margin) provides intuitive information how the targets’ motion varied throughout the run. The results obtained in both

figures (3a and 3b) unanimously show, as the motion speed of the targets increases, the performance difference between the two tracker modes widens. For example, at an approximate average speed of 1.25 m/s, the GPU-HOG-BIPBoost based tracker shows close to 20% gain in MOTA, compared to the less than 5% gain at around 0.5 m/s. The gap in tracker precision also widens similarly.

(2) *Safe Robotic Navigation*: in this evaluation, the robot is made to navigate from one starting point to an end point – which defines a mission. On each run, people are made to interfere with it by crossing its motion direction. If the robot manages to adjust its path without violating the person space around each person, then the mission is a success, otherwise it is a failure. This success/failure rate is quantified over several runs. Given our experimental environment is a bit narrow, a $4m \times 8m$ area, we have used a personal space with a major elliptical axis, a , of 1.0m and minor elliptical axis, b , of 0.6m (cf. Fig. 2). In exactly half of these experiments, the GPU-HOG is used as people detector, and for the other half, the GPU-HOG-BIPBoost is used. When using GPU-HOG-BIPBoost detector variant, the robot managed to complete missions successfully in 83% of the runs (out of 12 runs). The minimum distance encountered to a person is 74.4cm. All these values are superior to the alternative variant based on the GPU-HOG which achieves a 66.7% success rate, with an encountered minimum distance to a person of 67.2cm. These experimental results attest that the increased frame rate exhibited by the GPU-HOG-BIPBoost improves the robot’s reactivity.

IV. CONCLUSIONS

In this work, we demonstrate and quantitatively show the performance gain by using a faster visual people detector in a robotic dynamic obstacle avoidance application. The integrated GPU based BIPBoost detector improves the overall perceptual frame rate leading to improved people tracking performance – as demonstrated via on-line experiments – which in turn improve the reactivity of the mobile robot during navigation in human occupied environments. Specifically, utilizing this specific detector lead to an 83% robotic mission success rate in 12 runs.

REFERENCES

- [1] R. Benenson, M. Mathias, R. Timofte, and L. Van Gool, “Pedestrian detection at 100 frames per second,” in *IEEE CVPR*, 2012.
- [2] N. Dalal and B. Triggs, “Histograms of oriented gradients for human detection,” in *IEEE CVPR*, 2005.
- [3] A. A. Mekonnen, C. Briand, F. Lerasle, and A. Herbulot, “Fast HOG based person detection devoted to a mobile robot with a spherical camera,” in *IEEE/RSJ IROS*, 2013.
- [4] Z. Khan, T. Balch, and T. Dellaert, “Mcmc-based particle filtering for tracking a variable number of interacting targets,” *IEEE T-PAMI*, vol. 27, no. 11, pp. 1805–1918, 2005.
- [5] D. Vasquez, P. Stein, and J. Rios-Martinez et al., “Human Aware Navigation for Assistive Robotics,” in *ISER*, 2012.
- [6] J. Minguez and L. Montano, “Nearness diagram (ND) navigation: Collision avoidance in troublesome scenarios,” *IEEE T-RO*, vol. 20, no. 1, pp. 45 – 59, February, 2004.
- [7] K. Bernardin and R. Stiefelham, “Evaluating multiple object tracking performance: the CLEAR MOT metrics,” *EURASIP Journal on Image and Video Processing*, vol. 2008, pp. 1:1–1:10, January 2008.