A step towards human-aware path planning based on a combination of partial motion flows

Grimaldo Silva[1], Khansa Rekik[2], Ali Kanso[2]

***Abstract*— Automated analysis of people motion, particularly motion flow estimation, is of great importance to mobile navigation. Robots should be able to foresee the future motion of people and adjust their path accordingly in order to avoid situations with high risk of collision. This work reports on the progress towards an approach combining information from external video camera frames, what we call partial flows, to estimate people motion around the environment - a step towards human-aware path planning. Preliminary validation is done in a simplified simulated environment.**

I. INTRODUCTION

The use of mobile robots in both indoor and outdoor environments, such as in airports, malls, offices, etc., has been increasing over the last few years. Such has been also the case for the presence of cameras in such environments. The integration of both, however, has not yet been completely addressed by the literature.

Combining external cameras with mobile robots is advantageous in bringing a different level of awareness to robots during navigation. A robot could, for example, plan paths that avoid high density areas before any person has even been detected by its own sensors.

Unlike most state-of-the art methods that rely exclusively on robot internal sensors, this work adds partial external information about how the environment is populated wherever camera coverage is available. In a more general sense, our approach focuses on extracting information about the flow of people from each camera placed in the environment to estimate future human motion. An example is shown in Fig. 1. Preliminary validation is done in simulated experiments.

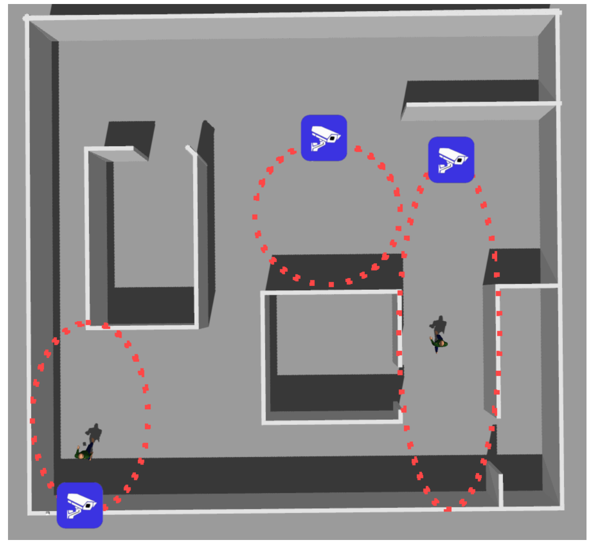


Fig. 1: Two out of three cameras positioned on the environment capture the motion of two people. Although more people can be hidden outside camera view, this "partial" observation of the flow of people can still allow for human-aware robot path planning with better situational awareness.

A. State of the art

From static or mobile platforms, at a local or a global level, people motion analysis has been investigated in literature for various reasons including, behavior understanding, event detection, motion detection etc. Surveys about crowd analysis and estimation techniques are available in [1], [2].

Towards motion prediction, both [3], [4] have used Long-Short Term Memory (LSTM) that enable learning human motion behavior from demonstrated data, for crowd motion prediction in cluttered environments. Combining crowd analysis and path planning, in [5] a technique was presented for navigation in dense crowds that combines trajectory prediction with Deep Reinforcement Learning-based collision avoidance. In [6], a method is described for environment feature extraction along with Inverse Reinforcement Learning (IRL) for socially adaptive navigation and path planning, while [7] introduced a social navigation system for human-populated and interactive scenarios using a method for clustering humans. However, these methods rely solely on robot internal sensors. Moreover, some have focused on cases where the observability of the environment is limited; [8] addressed the problem of inferring the human occupancy in the space around the robot *i.e.* blind spots, whereas [9] used IRL for path planning when limited knowledge of the density and flow of people in the environment is available. Our approach relies on using external cameras embedded in the environment to predict motion of people, providing additional situational awareness for future robot decision making.

In [10], smart video surveillance systems were investigated with a focus on people tracking from aerial video frames, crowd granularity analysis, and group event identification. While their method established a correspondence between consecutive frames. Our work focuses on estimating the partial motion flow on distinct regions and establishing a temporal relationship among them.

[1] Lactec, Parque Tecnológico da Bahia, 41745715, SSA, Brazil
[2] ZeMA gGmbH, Gewerbepark, Eschbergerweg 46, 66121 Saarbrücken
E-mail: jose.jgrimaldo@gmail.com, {k.rekik, a.kanso}@zema.de

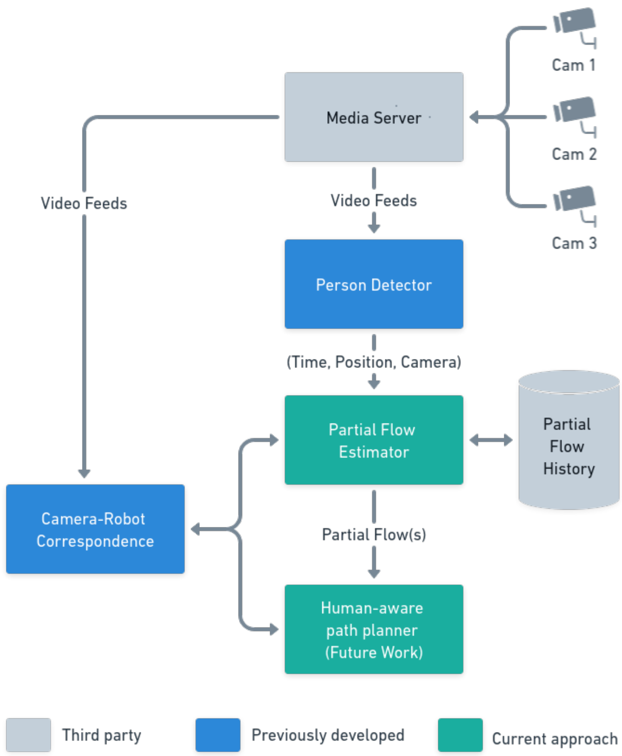


Fig. 2: Architecture of modules being developed for our approach and main components for its functionality. The "camera-robot" correspondence [11], autonomously map all camera pixels to the robot global coordinate space.

II. Partial motion flow representation

In order to achieve a human-aware robot navigation, it is essential to analyze the flow of people occurring in the areas the robot might visit. For that, videos from multiple cameras embedded in the environments are used. Each camera covers a region of the environment, we call the traffic in that area a partial flow F. To calibrate these cameras, a method for mapping camera images to robot world model is used.

A. Autonomous mapping of camera to robot frame

In our previous work [11] (under review)[1], a concept for controlling a robot directly through camera video feeds has been proposed. It is based on an approach for autonomous camera calibration based on the transformation between image pixels and robot coordinates using homography and deep learning. In other words, homographies for a bidirectional correspondence between camera space and the global coordinate space of the robot are autonomously calculated.

The concept was implemented and validated in simulated environments as well as real world conditions where the center of the robot was detected in the camera images using a deep learning robot detector, then associated with the robot position on its global coordinate to perform the calibration.

[1]For demonstration of [11], see `https://youtu.be/rZ1yFMNobSQ`

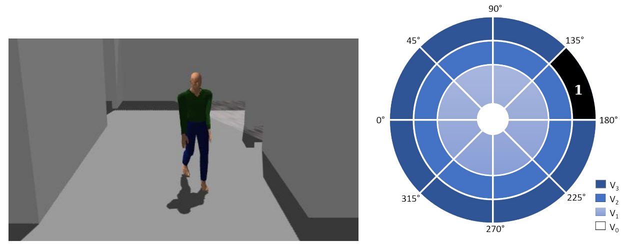


Fig. 3: A person moving towards the camera (left) with its partial flow captured and represented (right) as the black tile with the number of people with that particular flow type inside. Orientation is in the robot global coordinate space not the camera space. Let $V_0 \approx 0$ and V_1, V_2, V_3 respectively represent $0.25, 0.8, 1.35$ meters per second.

B. Feature space for partial flow

A partial flow is a combination of the previously observed flow in nearby cameras and the flow that is not temporally related to the appearance of people in other cameras. Each flow is represented by a feature vector f_c encoding number of people and their distribution of speed and direction.

In contrast to [6] who deals with people in terms of density and prioritize their relative position with respect to the robot, our approach focuses rather on the number of people and their motion inside regions where camera coverage is available, thus, the immediate neighborhood of the robot is not as relevant in this step. More specifically, we define 8 direction intervals ranging from 0° to 360° and 3 velocity intervals V_1, V_2 and V_3 ranging from the maximum velocity observed to $V_0 \approx 0m/s$ when the person is standing still. That yields in a $8 \times 3 + 1$ dimensional feature vector for each camera flow represented as shown in Fig. 3 (right).

III. Predicting near future partial flows

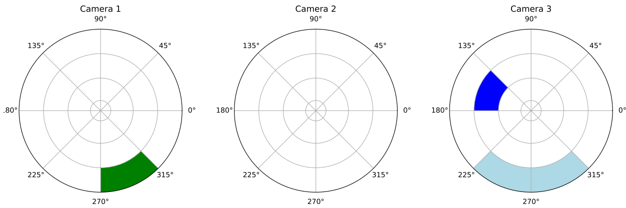
Consider a robot that wishes to efficiently translate from its current position to a goal position while accounting for people's motion in the environment. The robot is able to assess current partial flows calculated from every camera connected to the media server. However, as the robot plans a path and proceeds to execute it, the flow of people could already have changed.

As such, our objective is to estimate how the current flow of people will be in the future, based solely on current flows, after t seconds have elapsed.

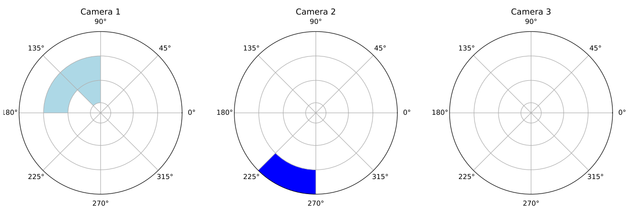
A. Partial Flow History

As shown in Fig. 2, the partial flows for each camera are continuously being saved. In order to avoid large differences in the number of samples per camera, due to distinct frames per second and other factors, flows are saved once every 2.5 seconds whenever a person has been detected in any camera for more than three consecutive frames in the last 30 seconds.

Although date information is not currently included in our experiments, usage in real world conditions would benefit from a time series representation in order to more directly represent reoccurring periods, such as: holidays, weekends and business days.



(a) Input feature vector with partial flows



(b) Output feature vector with regressor.

Fig. 4: Comparison between input feature vectors and output predictions. Each color indicates a person not feature value.

B. *Learning from observations*

Let C represent a set of cameras connected to the media server and $c_i \in C$ the camera of interest. The partial flow is calculated, in practice, in order to mitigate the impact of noise, using the mean partial flow observed over a time interval $[t-m;t]$ where $t \in \mathbb{N}$ is the current time and $m \in \mathbb{N}^+$ the interval for observation.

Our objective is to estimate, for c_i, the mean partial flow over of n seconds in the future as if observed over m seconds. Formally, the predicted value should be in the interval $[t+n;t+n+m]$ where $n \in \mathbb{N}$. This work assumes $n = 10$ and $m = 20$.

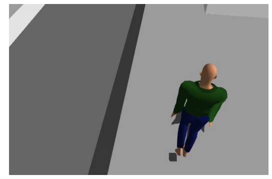
The mean partial flow history from every camera $c \in C$ are used as predictors to estimate the future flow of people in camera c_i in the future interval. For such prediction, in our simplified simulated scenario, an ordinary linear regressor[2] is tentatively used for training, where the to-be-predicted value is calculated from the partial flow history.

IV. PRELIMINARY SIMULATED VALIDATION

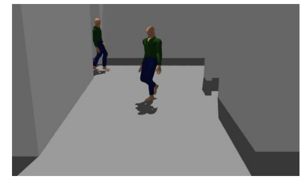
This section evaluates partial flow estimation in a simplified scenario. For our tests, a total of three people are set to travel the environment in pre-determined trajectories that are individually reset and repeat themselves ten seconds after a person arrives at their goal. As the time for each person to complete their trajectory is different, several distinct combinations of people appearing at each camera can be observed. Examples of people moving through the cameras in order to calculate partial flows are shown in Fig. 5

Preliminary results of the partial flow predictor are shown in Fig. 4, where, it successfully predicted future partial flows using current partial flows as predictors, however, for conclusive evidence more stochastic simulations are required.

[2]Deep learning regression is being considered for learning from timeseries-based real world experiments with people.



(a) Camera 1 where a person walks outside any camera view



(b) Camera 3, one person walks to camera 2 and another to 1

Fig. 5: Viewpoint of the cameras on our experimental setup

V. CONCLUSIONS

This work explored the concept of combining video streams from in world cameras to assist on the estimation of current and future flow of people in the environment. This work is only partially complete and still under development. Our current focus is on a more quantitative experimental validation and designing extensive real world experiments.

The objective of obtaining current and future partial flow is using it to build a human-aware costmap for path planning in a future work. As regions that lack camera coverage would not have an associated cost, an undesirable outcome, the challenge becomes estimating, from just partial flows, the motion of people throughout the whole environment.

References

[1] S. Saleh, S. Suandi, and H. Ibrahim, "Recent survey on crowd density estimation and counting for visual surveillance," *Engineering Applications of Artificial Intelligence*, vol. 41, pp. 103–114, 2015.

[2] H. Chaudhry, M. Rahim, T. Saba, and A. Rehman, "Crowd detection and counting using a static and dynamic platform: state of the art," *International Journal of Computational Vision and Robotics*, vol. 9, no. 3, pp. 228–259, 2019.

[3] A. Alahi, K. Goel, V. Ramanathan, A. Robicquet, L. Fei-Fei, and S. Savarese, "Social lstm: Human trajectory prediction in crowded spaces," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 961–971.

[4] M. Pfeiffer, G. Paolo, H. Sommer, J. Nieto, R. Siegwart, and C. Cadena, "A data-driven model for interaction-aware pedestrian motion prediction in object cluttered environments," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2018, pp. 5921–5928.

[5] A. Sathyamoorthy, J. Liang, U. Patel, T. Guan, R. Chandra, and D. Manocha, "Densecavoid: Real-time navigation in dense crowds using anticipatory behaviors," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2020, pp. 11 345–11 352.

[6] B. Kim and J. Pineau, "Socially adaptive path planning in human environments using inverse reinforcement learning," *International Journal of Social Robotics*, vol. 8, no. 1, pp. 51–66, 2016.

[7] A. Vega, L. J. Manso, D. G. Macharet, P. Bustos, and P. Núñez, "Socially aware robot navigation system in human-populated and interactive environments based on an adaptive spatial density function and space affordances," *Pattern Recognition Letters*, vol. 118, pp. 72–84, 2019.

[8] J. Amirian, J.-B. Hayet, and J. Pettre, "What we see and what we don't see: Imputing occluded crowd structures from robot sensing," *arXiv preprint arXiv:2109.08494*, 2021.

[9] P. Henry, C. Vollmer, B. Ferris, and D. Fox, "Learning to navigate through crowded environments," in *2010 IEEE International Conference on Robotics and Automation*. IEEE, 2010, pp. 981–986.

[10] P. Climent Perez, "Smart video surveillance of pedestrians: fixed, aerial, and multi-camera methods," Ph.D. dissertation, Kingston University, 2016.

[11] G. Silva, K. Rekik, and A. Kanso, "Multi-perspective human robot interaction through an augmented video interface supported by deep learning [under review]," in *Int. Conf. on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 2022.