Robot Navigation in Dense Human Crowds
Technical Report

Neil Traft

University of British Columbia

Overview

In their 2010 IROS publication [**?**], Trautman and Krause develop a path planning algorithm that is safe and yet does not suffer from the "freezing robot problem" (FRP). Their method consists of a model of crowd interaction combined with a particle-based inference method to predict where the crowd (and the robot) should be at some time $t+1$ in the future. The idea is that if one can develop a reliable model of intelligent agents in a crowd, and include the robot as just another of those intelligent agents, then the predictions of the model yield the robot's future path.

Interacting Gaussian Processes

The crowd interaction model is a novel nonparametric statistical model based on Gaussian processes. The authors have named it *Interacting Gaussian Processes* (IGP). In IGP, the actions of all agents, including the robot, are taken as a random response to other agents' behavior. Their interaction is modeled as a joint distribution:

$$p(\mathbf{f}^{(R)}, \mathbf{f}|\mathbf{z}_{1:t})$$

where $\mathbf{f}^{(R)}$ is the robot's trajectory over T timesteps, $\mathbf{f}$ is the set of all human trajectories, and $\mathbf{z}_{1:t}$ is the set of all observations up to the current time point. For the purposes of this algorithm, observations of human and robot position are taken to be more or less perfect, since we are only trying to solve the navigation problem, not the awareness problem.

Each agent's trajectory is an independent sample from a Gaussian process. This would be only a simple Gaussian process yielding the same uncertainty explosion which leads to the FRP, but it is modified in two ways. First, goal information is given as a final "observation" at time T, resulting in the full set of observations $\mathbf{z}_{1:t,T}$. The robot's goal, $y_T^{(R)}$, is known and can be added with good confidence. The goals of other agents can be omitted or can be added with a large confidence interval, to encode how uncertain we are about the goal.

The second addition IGP makes to standard Gaussian processes is the inclusion of an "interaction potential" $\psi(\mathbf{f}^{(R)}, \mathbf{f})$. In essence, this potential grows very small whenever two agents i and j become very close at any time τ. This has the result that any set of paths where agents become too close is treated as very unlikely. There are two parameters in this potential, h and α, that must be tuned for the desired "safety distance" and "repelling force". For details, see the paper.

Importance Sampling

Now that we have a model, we only need to sample from it to produce $\mathbf{f}^{(R)}$. However, with the addition of the interaction potential ψ, our distributions at each time step are nonlinear and potentially multimodal. Thus, we can no longer sample from the distribution directly. We instead use the *importance sampling* technique which is widely used in particle filters. Each sample is weighted by the ratio of the IGP to the basic GP (i.e. the Gaussian distribution, without the interaction potential):

$$w_i = \frac{p_{IGP}}{p_{GP}} = \frac{p_{IGP}((\mathbf{f}^{(R)}, \mathbf{f})|\mathbf{z}_{1:t})}{\prod_{j=R}^n p((\mathbf{f}^{(j)})_i|\mathbf{z}_{1:t})}$$

where $(\mathbf{f}^{(j)})_i$ is a single sample from the trajectory of agent j.

Given this formulation for p_{IGP} and an appropriate weighting w_i for each sample, we can now find the ideal paths:

$$(\mathbf{f}^{(R)}, \mathbf{f})^* = \arg\max p_{IGP}$$

and take the robot's next position to be $\mathbf{f}_{t+1}^{(R)}$.

Current Status

Currently I have downloaded the dataset that will be used for my simulations and begun visualization of the data.

The Dataset

The dataset to be used is the BIWI Walking Pedestrians dataset from [**?**]. It has been downloaded from [**?**].

The dataset contains two annotated videos of pedestrian interaction captured from a bird's eye view. Only one of these was used in Trautman's paper, a sequence acquired from the top of the ETH main building, Zurich, in 2009. The film captures people entering to and exiting from the main entrance of the building, such that agents in the image have two main goals: either their goal is a point at the bottom of the image (the entrance), or they are emerging from the entrance and their goal is somewhere along the top of the image (the street).

The main annotations are a matrix where each row has the format:

```
[frame_number pedestrian_ID pos_x pos_z pos_y v_x v_z v_y ]
```

Thus for each frame t we have potentially multiple pedestrian observations $\text{pos}_t^{(i)}$, and this forms our observation at time t:

$$\mathbf{z}_t = \text{pos}_t^{(1:n)}$$

where one of the n pedestrians is chosen to represent the robot R.

The `pos_z` and `v_z` (direction perpendicular to the ground) are not used in the annotations. The velocities will not be used by the IGP algorithm altogether; its only observations are positions at each timestep.

The positions and velocities are in meters and were obtained with a homography matrix H, which is also provided with the annotations. To transform the positions back to image coordinates, it is necessary to apply the inverse homography transform:

$$^a\text{pos}_t^{(i)} = H_{ab}^{-1} \cdot {}^b\text{pos}_t^{(i)}$$

Next Steps

Since I will only be simulating the planning algorithm on annotated video, and not implementing it on a robot, I expect it to be more expedient to implement the project in Python. I will use the Python Imaging Library (PIL) to digest the video frames. I will use NumPy, SciPy, and possibly some accompanying statistical packages to implement the algorithm. I will be able to visualize the results using matplotlib.

The dataset used in [**?**] is the BIWI Walking Pedestrians dataset from [**?**]. It is available online from the ETHZ Computer Vision Lab Dataset page. The annotations I parse from this dataset will be used to form the observations $\mathbf{z}_{1:t}$ used in the planning algorithm.

I will choose particular time points at which to run IGP, and choose a particular pedestrian to represent the "robot". By evaluating IGP at this single time slice, and comparing it to the actual path taken by the pedestrian, we can evaluate the performance of the planner. We evaluate its performance based on l, the length of the path, and s, the minimum distance which the path ever comes within another pedestrian's path.

I expect to be able to closely reproduce the original results. I will have to tweak the parameters α and h in the interaction potential, since the ones used for the original experiments are not directly given. I will also have to tweak the number of particles used for inference, though the paper does specify a range of $[100, 5000]$ which should help. I may also have to play with the "simulated observation" that represents each agent's goal. The original paper does not give any details about the use of agent goals.

References

[1] P. Trautman and A. Krause, "Unfreezing the robot: Navigation in dense, interacting crowds," *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 797–803, Oct. 2010.

[2] S. Pellegrini, a. Ess, K. Schindler, and L. van Gool, "You'll never walk alone: Modeling social behavior for multi-target tracking," *2009 IEEE 12th International Conference on Computer Vision*, pp. 261–268, Sept. 2009.

[3] "Ethz - computer vision lab: Datasets." http://www.vision.ee.ethz.ch/datasets/index.en.html. Accessed: 2014-03-01.