Motivation

Humans navigate through crowds by adapting their trajectories to those of other people in the vicinity. Robots need to be able to predict trajectories of surrounding humans to move in a socially compliant fashion through crowds. However, current approaches cannot capture the complex and often subtle interactions among humans in crowd.

Objectives

- **Group behavior**
- **Collision avoidance**

Such interactions cannot be modeled using an independent prediction model:

$$ P(f_1, f_2, \ldots, f_n) = \prod_{i=1}^{n} P(f_i) $$  

We need to model the joint distribution of trajectories of all agents:

$$ P(f_1, f_2, \ldots, f_n) $$

The resulting model should:
- Scale well with number of agents
- Accurate long-term predictions

Results

- **IGP**
  - Better long-term predictions
  - Captures cooperative and collision avoidance behavior
  - Learns relevance of each surrounding agent to a pedestrian’s motion
  - Accurate long-term prediction results in a more socially compliant path prediction for the robot

Approach

Joint Prediction Model

Interacting Gaussian Processes (Trautman and Krause 2010):

$$ P(f_1, \ldots, f_n) = \frac{1}{Z} \psi(f_1, f_2, \ldots, f_n) \prod_{i=1}^{n} P(f_i) $$

Handcrafted potential

Our formulation using occupancy grids:

$$ P(f_1, \ldots, f_n) = \sum_{\text{grid}} \prod_{i=1}^{n} P(f_i|O_i, g_i) $$

Given observed part of trajectories ($z_i^{t_1:t_{t+1}}$):

$$ P(f_1, f_2, \ldots, f_n|z_i^{t_1:t_{t+1}}) = \sum_{\text{grid}} \prod_{i=1}^{n} P(f_i|O_i, g_i, z_i^{t_1:t_{t+1}}) P(g_i|z_i^{t_1:t_{t+1}}) $$

Learning and Inference

Learning from training data:

- Predict velocities and compute occupancy grids at each time-step

Inference:

- Use the learned model to infer goals from observed trajectory
- Predict velocities and compute occupancy grids at each time-step

avemula1@andrew.cmu.edu, {kmuelling, jeanoh}@nrec.ri.cmu.edu
Modeling Cooperative Navigation in Dense Human Crowds

Anirudh Vemula  Katharina Muelling  Jean Oh

Robotics Institute, Carnegie Mellon University

Overview

Objectives
• Model the joint distribution of trajectories of all agents
• Scale well with the number of agents
• Capture cooperative behavior in crowds

Approach
Using occupancy grids to capture local vicinity

Predicting future trajectories:

1. For each agent \( i \) do
2. Compute distributions of \( (v_i)^{t+1} \), \( (y_i)^{t+1} \) given \( O_t \)
3. end for
4. for \( t'=t+2 \rightarrow t+H \) do
5. for each agent \( i \) do
6. Sample \( S \) points from distributions of \( (v_i)^{t-1} \), \( (y_i)^{t-1} \)
7. Compute \( S \) estimates for \( f_i^{t} \) from sampled velocities
8. end for
9. Compute \( S \) samples for \( O_t' \) from estimates of \( f_t' \)
10. Set \( O_t' \) to be the mean of the \( S \) samples from above
11. for each agent \( i \) do
12. Compute distributions of \( (v_i)^{t'}, (y_i)^{t'} \) given \( O_t' \)
13. end for
14. end for

Results
• Better long-term predictions
• Captures cooperative and collision avoidance behavior

Constructing Occupancy Grids

Location

Occupancy grid is a vector of length \( M^2 \) capturing occupancy in the neighborhood of agent \( i \) discretized as a \( M \times M \) grid.

\[
\text{Occupancy grid} = \sum_{n=1}^{N} I_{n} \delta[x'_{n} - x_{i}, y'_{n} - y_{i}].
\]

\( n \) \text{Count of agents occupying this cell}

\( x_{i}, y_{i} \) \text{Surrounding agent coordinates}

INference

Inferring goal of agent: Given observations \( z_{i}^{t} \) of agent \( i \) until time \( t \),

\[
P(g_{i}|z_{i}^{t}) = \frac{P(z_{i}^{t}|g_{i})P(g_{i})}{P(z_{i}^{t})} \propto P(z_{i}^{t}|g_{i})
\]

Normalizing the marginal likelihoods of GPs across all goals, we get the likelihood \( P(z_{i}^{t}|g_{i}) \)

Predicting future trajectories: Now that we have a distribution over the goals \( g_{i} \) for all agents \( i \) in the crowd, we can use the trained model to predict future locations. The joint posterior density can be decomposed as

\[
P(f_{1}, \ldots, f_{n}, |g_{i}^{t}, z_{i}^{t+1}) = \prod_{g_{i}} P(f_{i}^{t}|g_{i}) P(g_{i}^{t})
\]

Assuming goals of each agent is independent of each other we have,

\[
P(g_{i}^{t}|z_{i}^{t+1}) = \prod_{n=1}^{N} P(g_{i}^{t}|z_{i}^{t+1})
\]

We approximate the joint distribution as

\[
P(f_{1}, \ldots, f_{i}|g_{i}^{t}, z_{i}^{t+1}) = P(f_{1}, \ldots, f_{i}||v_{i}, v_{y1}, t_{i+1}, O_{i}^{t+1}, g_{i}^{t})
\]
Modeling Cooperative Navigation in Dense Human Crowds
Anirudh Vemula  Katharina Muelling  Jean Oh
Robotics Institute, Carnegie Mellon University

**Overview**

**Objectives**
- Model the joint distribution of trajectories of all agents
- Scale well with the number of agents
- Capture cooperative behavior in crowds

**Approach**

Using occupancy grids to capture local vicinity

\[
P(f_1, \ldots, f_n) = \sum_g \prod_{i=1}^n P(f_i | O_i, g)
\]

Observed trajectory
Predicted trajectory
Predict velocities and compute occupancy grids at each time-step

**Results**

- Better long-term predictions
- Captures cooperative and collision avoidance behavior

**Collision Avoidance**

Given sample occupancy grids, our model predicts velocities that avoid collisions with surrounding agents, and also exhibits behaviors such as slowing down.

**Automatic Relevance Determination**

**Results**

Dataset with goals marked
Cooperative behavior

<table>
<thead>
<tr>
<th>Metric</th>
<th>Prediction Horizon (t)</th>
<th>IGP</th>
<th>Our Approach</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avg. Disp. Error</td>
<td>1</td>
<td>3.42</td>
<td>4.42</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>5.66</td>
<td>6.14</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>15.75</td>
<td>12.09</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>21.59</td>
<td>21.52</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>41.51</td>
<td>34.63</td>
</tr>
<tr>
<td>Final Disp. Error</td>
<td>1</td>
<td>3.42</td>
<td>4.42</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>7.12</td>
<td>7.78</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>23.18</td>
<td>19.77</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>38.75</td>
<td>36.25</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>67.41</td>
<td>54.2</td>
</tr>
</tbody>
</table>

Each pixel in the video frame corresponds to 0.042 metres (slightly varies across the frame, as the camera is angled and not exactly top-down)

avemula1@andrew.cmu.edu, {kmuelling, jeanoh}@nrec.ri.cmu.edu