Tracking in Dense Crowds

Goal: Track targets in high-density crowd scenes.

Challenges: lots of occlusion; small object sizes; appearances are similar

Idea: Model typical crowd behavior to provide better motion priors.
Point of View: Macro vs Micro

• Macroscopic level: modeling dynamic behavior of the whole crowd; holistic
  – density, flow, mean speed of a traffic stream
  – analogy to fluid streams; particle flow
  – behavior is reactive, a function of environment and density

  Crowd Flow

• Microscopic level: models decision makers, their goals, and interactions; individualistic
  – intelligent agents make decisions based on goals and social rules
  – simulating realistic interactions

  Social Force Models
Crowd Flow: Floor Fields


Inspired by particle flow evacuation models.

Represents how global scene structure affects local pedestrian motion decisions.

Long-range goals/influences transformed into local forces (similar to potential fields for robotic path planning).
Floor Fields

• **Static Floor Field (SFF)**
  attraction field; represents typical crowd motion towards interesting locations, dominant paths, exits

• **Boundary Floor Field (BFF)**
  repulsive forces; boundaries, walls, obstacles

• **Dynamic Floor Field (DFF)**
  current motion of neighboring individuals computed in temporal sliding window
Static Floor Field

example: marathon runners turning a corner

optic flow  averaged flow over time  sink-seeking

mean-shift-like procedure to determine particle flow (path, distance) to nearest goal location.
Static Floor Field

example: marathon runners turning a corner

optic flow  averaged flow over time  sink-seeking

SFF = path length surface. Low values are “better”. Intuition: drop a ball on surface and it rolls towards nearest sink.
Boundary Floor Field

averaged flow

segmented flow

down map
(real+virtual boundaries)

BFF = truncated distance transform. High values are “better”. Intuition: go/no-go surface with deep valleys forming the barriers.
Dynamic Floor Field

local neighborhood around target location (yellow dot)

DFF = current local motion likelihood computed from flow in a narrow temporal window.

Intuition: this is how nearby particles are currently moving.
How Floor Fields are Used

For current target location, compute matrix of local transition probabilities combining appearance and floor field terms.

\[ p_{ij} = Ce^{k_D D_{ij}}e^{k_S S_{ij}}e^{k_B B_{ij}}R_{ij} \]

- **SFF/BFF/DFF influence terms** (priors)
- **appearance term** (likelihood)
How Floor Fields are Used

multimodal likelihood (appearance is not discriminative)

local motion prior

scene goal prior

much more reliable (unimodal) posterior
Tracking Examples
Tracking Examples
Floor Field Drawbacks

• SFF can’t represent multimodal goals / motion at single point in the scene

• DFF allows some local temporal adaptation, but only correct when target moves similar to neighbors

• Hard to track outlier behaviors (moving against traffic)
HMM-based Flow Model


Intuition: model multi-modal, time-varying flow by training an HMM at each scene location.
HMM-based Flow Model

Training stage:

- **dice training video into space-time cuboids**
- **estimate 3D Gaussian motion pattern in each cuboid** (space-time gradients)
- **in each time-tube of cuboids**
  - discretize motion patterns by online clustering
  - train an HMM
HMM-based Flow Model

Training stage:

The HMMs can model time-dependencies between multiple motions at a single spatial location.

e.g. “this location has two dominant flow directions that tend to be interleaved”
    “this location exhibits many rapidly-changing flow directions”
    “this location has a single dominant flow”
HMM-based Flow Model

Tracking stage:

- at runtime, use observed motion patterns up to time $t-1$ to compute expected motion at target’s center at time $t$.

- project this 3D motion pattern into 2D to get predicted image flow distribution

- use this distribution as a motion prior for particle filter tracking
Sample Results

play video outside ppt
Data Driven Flow Modeling

• Floor fields and HMM-based flow are scene-centric models (must be trained previously on video from the same scene viewpoint)

• They also have trouble tracking “rare” motions because they accumulate distributions of typical scene behavior

• Idea: try non-parametric data-driven approaches that have been very successful in texture synthesis and inpainting.
Data-Driven Flow


Insight: Any given crowd video can be viewed as a composite mixture of patches taken from a large dataset of previously viewed videos.
Two-Stage Matching

- First stage: Global matching using GIST descriptor of first frame to find videos roughly matching orientation and scale (viewpoint) of input video.

input video matches from database
Two-Stage Matching

- Second stage: Local patch matching based on HOG3D descriptors (histograms of spatio-temporal gradients) to find patches with similar structure and motion as neighborhood around target.
Motion Transfer

- Motion information is averaged over the matching patches and incorporated into a motion prior during Kalman filter tracking.
- This data-driven prior, using different videos, does better than averaging scene flow over the actual input sequence.

red = ground truth; green = data-driven flow, yellow = averaged scene flow
Performance on Rare Events

Figure 9. Comparison of average tracking errors when tracking people in rare crowd events based on 21 tracks and $k = 3$. 
Social Force Model


Social forces represent similar information as floor fields.

But one important distinction: working in an agent-centered point of view rather than a scene-centered one.

In other words, microscopic rather than macroscopic.
Social Force Model

you are here

desired velocity

goal
Social Force Model

you are here

repulsive force

obstacle

goal
Social Force Model

You are here

Something interesting

Attractive force

Obstacle

Goal

Robert Collins
Penn State

VLPR 2012
Social Force Model

You are here

Collision avoidance

Something interesting

Other pedestrians

Obstacle

Goal

VLPR 2012
Social Force Model

- you are here
- desired velocity
- obstacle
- something interesting
- other pedestrians
- goal
Social Force Model

you are here

you are here

actual velocity

desired velocity

obstacle

something interesting

other pedestrians

game
Consider two moving pedestrians.

What is their point of closest approach?
(assuming they move with constant velocity)
Case Study


\[
p_1(t) = s_1 + t \, v_1 \quad \quad p_2(t) = s_2 + t \, v_2
\]

\[
t^* = \arg\min_{t > 0} || p_2(t) - p_1(t) ||
\]

\[
c_1 = s_1 + t^* \, v_1 \quad \quad c_2 = s_2 + t^* \, v_2
\]
intuition: we want to adjust $v_1$ and $v_2$ to keep a "comfortable" distance $d_{12}$ between them, while maintaining roughly the original desired directions and speeds.
the new velocities are found by numerical optimization

s1 slows down and turns to right

s2 speeds up and also turns to right
Model Yields Intuitive Behavior


Depending on distance between $s_2$ and $s_3$, pedestrian $s_1$ will either try to pass between them, or around them.
Pedestrian Fingering

Helbing’s social force model also predicts “fingering” in areas of bidirectional motion. People tend to follow others to minimize collisions (maximize throughput).

Green: leftward moving. Red: rightward moving,
leftrightMovie.mp4
Fingering Effect

collective behavior emerges from independent decisions

Density by image row
Green (leftward); Red (rightward)
Collective Locomotion

• Find small groups traveling together
  – Sociological hypothesis: validating that the majority of people in the crowd cluster in small groups
  – Public safety: improving situation awareness and emergency response during public disturbances
McPhail and Wohlstein, 1982

- Group membership is determined via a cascaded set of three tests:
  1. Any two people who are within 7 feet of each other and not separated by another individual are considered to be contiguous.
  2. Any two contiguous people whose speeds are the same to within .5 feet per second are judged to have the same speed.
  3. Any two contiguous people traveling at the same speed whose directions of motion are the same to within 3 degrees are judged to have the same direction.

- Another procedure tests whether a new individual should be added to an existing group to form a larger group.

- Limitations
  - Hundreds of person-hours needed to hand code just minutes of film.
  - Difficult for dense crowds/long sequences.

Automated Group Testing by Agglomerative Clustering


"distance" is based on spatial proximity and velocity coherence.
Sample Results

Evaluation reveals substantial agreement between computer-generated groupings and those found by human coders (ground truth)

<table>
<thead>
<tr>
<th></th>
<th>match rate</th>
<th>$\chi^2(4, 248)$</th>
<th>Cohen’s $\kappa$</th>
</tr>
</thead>
<tbody>
<tr>
<td>trichotomous</td>
<td>85%</td>
<td>219.98</td>
<td>.69</td>
</tr>
<tr>
<td>dichotomous</td>
<td>89%</td>
<td>138.26</td>
<td>.75</td>
</tr>
</tbody>
</table>

$p < .001$
More Grouping Results
Likely Group Shapes

Are some group configurations more likely than others? Of course!
Analysis of Group Shape

Figure 5.14. The configurations of groups of size three are aligned with respect to their group centers and moving directions. The three members are plotted with three different colors after a data association procedure that matches points across different configurations. Edges indicating the group configuration are omitted for clarity.
Analysis of Group Shape

Procrustes Analysis, first four modes of variation

video3PersonGroupShape.mp4
Research Questions

Is multitarget tracking of human crowds any different than tracking crowds of animals? bats? cells?