Combining 3D Shape, Color, and Motion for Robust Anytime Tracking
David Held, Jesse Levinson, Sebastian Thrun, and Silvio Savarese

Goal: Fast and Robust Velocity Estimation

Our Approach: Alignment Probability
- Spatial Distance
- Color Distance (if available)
- Probability of Occlusion

Annealed Dynamic Histograms

\[
p(x_t \mid z_1 \ldots z_t) = \eta p(z_t \mid x_t, z_{t-1})p(x_t \mid z_1 \ldots z_{t-1})
\]
Combining 3D Shape, Color, and Motion for Robust Anytime Tracking

David Held, Jesse Levinson, Sebastian Thrun, and Silvio Savarese

Goal: Fast and Robust Velocity Estimation

Baseline: Centroid Kalman Filter

Baseline: ICP

Local Search → Poor Local Optimum!

Baseline: Centroid Kalman Filter

Baseline: ICP

Baseline: Fast and Robust Velocity Estimation

Combining 3D Shape, Color, and Motion for Robust Anytime Tracking

David Held, Jesse Levinson, Sebastian Thrun, and Silvio Savarese

Goal: Fast and Robust Velocity Estimation

Baseline: Centroid Kalman Filter

Baseline: ICP

Local Search → Poor Local Optimum!

Baseline: Fast and Robust Velocity Estimation

Combining 3D Shape, Color, and Motion for Robust Anytime Tracking

David Held, Jesse Levinson, Sebastian Thrun, and Silvio Savarese

Goal: Fast and Robust Velocity Estimation

Baseline: Centroid Kalman Filter

Baseline: ICP

Local Search → Poor Local Optimum!

Baseline: Fast and Robust Velocity Estimation

Combining 3D Shape, Color, and Motion for Robust Anytime Tracking

David Held, Jesse Levinson, Sebastian Thrun, and Silvio Savarese

Baseline: Centroid Kalman Filter

Baseline: ICP

Local Search → Poor Local Optimum!

Baseline: Fast and Robust Velocity Estimation

Combining 3D Shape, Color, and Motion for Robust Anytime Tracking

David Held, Jesse Levinson, Sebastian Thrun, and Silvio Savarese

Baseline: Centroid Kalman Filter

Baseline: ICP

Local Search → Poor Local Optimum!

Baseline: Fast and Robust Velocity Estimation

Combining 3D Shape, Color, and Motion for Robust Anytime Tracking

David Held, Jesse Levinson, Sebastian Thrun, and Silvio Savarese

Baseline: Centroid Kalman Filter

Baseline: ICP

Local Search → Poor Local Optimum!

Baseline: Fast and Robust Velocity Estimation

Combining 3D Shape, Color, and Motion for Robust Anytime Tracking

David Held, Jesse Levinson, Sebastian Thrun, and Silvio Savarese

Baseline: Centroid Kalman Filter

Baseline: ICP

Local Search → Poor Local Optimum!
Combining 3D Shape, Color, and Motion for Robust Anytime Tracking

David Held, Jesse Levinson, Sebastian Thrun, and Silvio Savarese

Goal: Fast and Robust Velocity Estimation

Our Approach: Alignment Probability

- Spatial Distance
- Color Distance (if available)
- Probability of Occlusion

Annealed Dynamic Histograms

\[
p(x_t \mid z_1 \ldots z_t) = \eta p(z_t \mid x_t, z_{t-1})p(x_t \mid z_1 \ldots z_{t-1})
\]
Motivation

Quickly and robustly estimate the speed of nearby objects
System

Camera Images

Laser Data
System

Camera Images

Laser Data

Previous Work
(Teichman, et al)
System

Camera Images

Laser Data

Previous Work (Teichman, et al)

This Work

Velocity Estimation
Velocity Estimation
Velocity Estimation
Velocity Estimation
Velocity Estimation
Velocity Estimation

Occlusions  Actual Motion
ICP Baseline
ICP Baseline
ICP Baseline
ICP Baseline
ICP Baseline

Local Search → Poor Local Optimum!
Tracking Probability
Velocity Estimation
Velocity Estimation

t+1  

t
Velocity Estimation
Velocity Estimation
Velocity Estimation
Velocity Estimation
Velocity Estimation

\[ X_t \]

\[ t \]

\[ t+1 \]
Velocity Estimation

\[ p(x_t \mid z_1 \ldots z_t) \]
Tracking Probability

\[ p(x_t \mid z_1 \ldots z_t) = \eta p(z_t \mid x_t, z_{t-1}) p(x_t \mid z_1 \ldots z_{t-1}) \]

Measurement Model

Motion Model
Tracking Probability

\[ p(x_t \mid z_1 \ldots z_t) = \eta p(z_t \mid x_t, z_{t-1}) p(x_t \mid z_1 \ldots z_{t-1}) \]

- Measurement Model
- Motion Model
- Constant velocity Kalman filter
Tracking Probability

\[ p(x_t \mid z_1 \ldots z_t) = \eta p(z_t \mid x_t, z_{t-1}) p(x_t \mid z_1 \ldots z_{t-1}) \]

Measurement Model

Motion Model
Tracking Probability

\[
p(x_t \mid z_1 \ldots z_t) = \eta \, p(z_t \mid x_t, z_{t-1}) \, p(x_t \mid z_1 \ldots z_{t-1})
\]
Tracking Probability

\[ p(x_t \mid z_1 \ldots z_t) = \eta p(z_t \mid x_t, z_{t-1}) p(x_t \mid z_1 \ldots z_{t-1}) \]

Measurement Model

Motion Model
Tracking Probability

\[
p(x_t \mid z_1 \ldots z_t) = \eta p(z_t \mid x_t, z_{t-1}) p(x_t \mid z_1 \ldots z_{t-1})
\]
Tracking Probability

\[
p(x_t \mid z_1 \ldots z_t) = \eta p(z_t \mid x_t, z_{t-1}) p(x_t \mid z_1 \ldots z_{t-1})
\]
Tracking Probability

\[ p(x_t \mid z_1 \ldots z_t) = \eta p(z_t \mid x_t, z_{t-1}) p(x_t \mid z_1 \ldots z_{t-1}) \]

Measurement Model

Motion Model
Tracking Probability

\[ p(x_t \mid z_1 \ldots z_t) = \eta p(z_t \mid x_t, z_{t-1}) p(x_t \mid z_1 \ldots z_{t-1}) \]

Measurement Model

Motion Model
Tracking Probability

\[ p(x_t \mid z_1 \ldots z_t) = \eta \, p(z_t \mid x_t, z_{t-1}) \, p(x_t \mid z_1 \ldots z_{t-1}) \]

Measurement Model

\[ = \eta \left( \prod_{z_i \in z_t} \exp \left( -\frac{1}{2} (z_i - \bar{z}_j)^T \Sigma^{-1} (z_i - \bar{z}_j) + k \right) \right) \]
\[
p(x_t \mid z_1 \ldots z_t) = \eta \frac{p(z_t \mid x_t, z_{t-1})}{\text{Measurement Model}} \frac{p(x_t \mid z_1 \ldots z_{t-1})}{\text{Motion Model}}
\]

\[
= \eta \left( \prod_{z_i \in z_t} \exp \left( -\frac{1}{2} \left( z_i - \bar{z}_j \right)^T \Sigma^{-1} \left( z_i - \bar{z}_j \right) \right) \right) + k
\]
\[ p(x_t \mid z_1 \ldots z_t) = \eta p(z_t \mid x_t, z_{t-1}) \cdot p(x_t \mid z_{1 \ldots t-1}) \]

\[ = \eta \left( \prod_{z_i \in z_t} \exp \left( -\frac{1}{2} (z_i - \bar{z}_j)^T \Sigma^{-1} (z_i - \bar{z}_j) + k \right) \right) \]
Tracking Probability

\[ p(x_t \mid z_1 \ldots z_t) = \eta p(z_t \mid x_t, z_{t-1}) p(x_t \mid z_1 \ldots z_{t-1}) \]

\[ = \eta \left( \prod_{z_i \in z_t} \exp \left( -\frac{1}{2} (z_i - \bar{z}_j)^T \Sigma^{-1}(z_i - \bar{z}_j) \right) + k \right) \]
\[ p(x_t \mid z_1 \ldots z_t) = \eta p(z_t \mid x_t, z_{t-1}) p(x_t \mid z_1 \ldots z_{t-1}) \]

Measurement Model

Motion Model

\[
\sum_e \sum_{s_{t-1}} \sum_{z_{t-1}} \sum_{e} \sum_{s_t} \sum_{z_t} = \eta \left( \prod_{z_i \in z_t} \exp \left( -\frac{1}{2} (z_i - \bar{z}_j)^T \Sigma^{-1} (z_i - \bar{z}_j) \right) + k \right)
\]
Tracking Probability

\[ p(x_t \mid z_1 \ldots z_t) = \eta \frac{p(z_t \mid x_t, z_{t-1})}{\sum_r} p(x_t \mid z_1 \ldots z_{t-1}) \]

Measurement Model

Motion Model

\[ \sum_e = 2\sum_e + \sum_r \]

Sensor noise

Sensor resolution

\[ \sum_e = \eta \left( \prod_{z_i \in z_t} \exp \left( -\frac{1}{2} (z_i - \bar{z}_j)^T \Sigma^{-1} (z_i - \bar{z}_j) \right) + k \right) \]
Color Probability

![Graph showing the relationship between Delta Color Value and Probability. The x-axis represents Delta Color Value, ranging from 0 to 250. The y-axis represents Probability, ranging from 0 to 0.08. The graph shows a decreasing trend as Delta Color Value increases.]
Including Color

\[ p(z_t \mid x_t, z_{t-1}) = \eta \prod_{z_i \in z_t} p_s(z_i \mid \bar{z}_j) \]
Including Color

\[ p(z_t \mid x_t, z_{t-1}) = \eta \prod_{z_i \in z_t} p_s(z_i \mid \bar{z}_j) \]

\[ \eta \left( \prod_{z_i \in z_t} \exp(-\frac{1}{2}(z_i - \bar{z}_j)^T \Sigma^{-1}(z_i - \bar{z}_j)) + k \right) \]
Including Color

\[ p(z_t \mid x_t, z_{t-1}) = \eta \prod_{z_i \in z_t} p_s(z_i \mid \bar{z}_j)p_c(z_i \mid \bar{z}_j) \]

\[ \eta \left( \prod_{z_i \in z_t} \exp\left(-\frac{1}{2}(z_i - \bar{z}_j)^T \Sigma^{-1}(z_i - \bar{z}_j)\right) + k \right) \]
Including Color

\[ p(z_t \mid x_t, z_{t-1}) = \eta \prod_{i \in z_t} p_s(z_i \mid \bar{z}_j)p_c(z_i \mid \bar{z}_j) \]

\[ p(C)p(z_t \mid \bar{z}_j, C) + P(\neg C)p(z_t \mid \bar{z}_j, \neg C) \]
Including Color

\[ p(z_t \mid x_t, z_{t-1}) = \eta \prod_{z_i \in z_t} p_s(z_i \mid \bar{z}_j) p_c(z_i \mid \bar{z}_j) \]

\[ p(C)p(z_t \mid \bar{z}_j, C) + P(\neg C)p(z_t \mid \bar{z}_j, \neg C) \]

\[ p_c \exp\left(\frac{-r^2}{2\sigma_c^2}\right) \]
Including Color

\[ p(z_t \mid x_t, z_{t-1}) = \eta \prod_{z_i \in z_t} p_s(z_i \mid \bar{z}_j) p_c(z_i \mid \bar{z}_j) \]

\[ p(C)p(z_t \mid \bar{z}_j, C) + P(\neg C)p(z_t \mid \bar{z}_j, \neg C) \]

\[ p_c \exp\left(\frac{-r^2}{2\sigma_C^2}\right) \]
Including Color

\[ p(z_t \mid x_t, z_{t-1}) = \eta \prod_{z_i \in z_t} p_s(z_i \mid \bar{z}_j) p_c(z_i \mid \bar{z}_j) \]

\[ p(C)p(z_t \mid \bar{z}_j, C) + P(\neg C)p(z_t \mid \bar{z}_j, \neg C) \]

\[ p_c \exp\left(\frac{-r^2}{2\sigma^2_c}\right) \]

\[ 1 - p_c \exp\left(\frac{-r^2}{2\sigma^2_c}\right) \]
Including Color

\[ p(z_t \mid x_t, z_{t-1}) = \eta \prod_{z_i \in z_t} p_s(z_i \mid \bar{z}_j) p_c(z_i \mid \bar{z}_j) \]

\[ p(C)p(z_t \mid \bar{z}_j, C) + P(\neg C)p(z_t \mid \bar{z}_j, \neg C) \]

\[ p_c \exp\left(\frac{-r^2}{2\sigma^2_c}\right) \]

\[ 1 - p_c \exp\left(\frac{-r^2}{2\sigma^2_c}\right) \]

\[ \frac{1}{255} \]
Probabilistic Framework

3D Shape

Color

Motion History

Tracking
Tracking Probability

- $P_1$
- $P_2$
- $P_3$
- $P_4$
Tracking Probability

? ?
? ?
? ?

$v_y$

$v_x$
Tracking Probability
Dynamic Decomposition
Dynamic Decomposition
Dynamic Decomposition
Dynamic Decomposition

Derived from minimizing KL-divergence between approximate distribution and true posterior
Annealing

Inflate the measurement model
Annealing

Inflate the measurement model
Annealing

Inflate the measurement model
Algorithm

1. For each hypothesis

   A. Compute the probability of the alignment

\[
p(x_t \mid z_1 \ldots z_t) = \eta p(z_t \mid x_t, z_{t-1}) p(x_t \mid z_1 \ldots z_{t-1})
\]

Measurement Model  Motion Model
Algorithm

1. For each hypothesis
   A. Compute the probability of the alignment
      \[ p(x_t \mid z_1 \ldots z_t) = \eta p(z_t \mid x_t, z_{t-1}) p(x_t \mid z_1 \ldots z_{t-1}) \]
      Measurement Model
      Motion Model
   B. Finely sample high probability regions
Algorithm

1. For each hypothesis
   A. Compute the probability of the alignment
      \[ p(x_t \mid z_1 \ldots z_t) = \eta p(z_t \mid x_t, z_{t-1}) p(x_t \mid z_1 \ldots z_{t-1}) \]
      Measurement Model
      Motion Model
   B. Finely sample high probability regions
   C. Go to step 1 to compute the probability of new hypotheses
Annealing

More time
More accurate
Anytime Tracker

![Graph showing the relationship between Mean runtime (ms) and RMS error (m/s). The graph indicates a decrease in RMS error as Mean runtime increases.]
Choose runtime based on:
- Total runtime requirements
- Importance of tracked object

...
Comparisons

![Comparison diagram showing RMS error (m/s) versus Mean runtime (ms) for different algorithms: Kalman Filter, ICP, Kalman ICP with Centroid Init, and Annealed Dynamic Histograms. The diagram indicates that Kalman Filter has a unique point with a higher RMS error and lower runtime compared to the other algorithms.]
Comparisons

![Comparison Graph]

- **Kalman Filter**
- **ICP**
- **Kalman ICP with Centroid Init**
- **Kalman ICP with Kalman Init**

RMS error (m/s) vs. Mean runtime (ms)
Comparisons

![Comparison graph showing RMS error (m/s) vs. Mean runtime (ms) for different algorithms. Key:
- X: Kalman Filter
- *: ICP
- •: Kalman ICP with Centroid Init
- ••: Kalman ICP with Kalman Init
- : Annealed Dynamic Histograms]
Kalman Filter
Models
Quantitative Evaluation 2

![Graph showing crispness for People, Bikes, and Cars with different methods: Kalman Filter, Kalman ICP, and ADH Tracker (Ours).]
Sampling Strategies

- ADH Tracker (Ours)
- Dense sampling
- Dense sampling with motion prediction
- Top cell sampling
Advantages over Radar
Conclusions

- Robust to Occlusions, Viewpoint Changes
Conclusions

- 3D Shape
- Color
- Motion History

- Robust to Occlusions, Viewpoint Changes
- Runs in Real-time
- Robust to Initialization Errors
Error vs Number of Points

RMS error (m/s)

Kalman Filter
ADH Tracker (Ours)

Number of points

0 200 400 600 800
Error vs Distance

- Kalman Filter
- ADH Tracker (Ours)
Error vs Number of Frames

RMS error (m/s) vs Number of frames tracked.
Error vs Number of Frames

- **RMS error (m/s)**
- **Number of frames tracked**

Graph showing the comparison between ADH Tracker and ADH Tracker without motion model.
The graph compares RMS error (m/s) for different conditions:

- No color
- No motion model
- No 3D shape

The error values are higher for the conditions involving no 3D shape compared to those with no color or no motion model.