Combining Local and Global Motion Models for Feature Point Tracking

Problem: determine the motion of a feature \( x_t \) given observations \( z_{t-1} \).

We employ a robust motion prior derived from the global motion in the scene that can be combined naturally with the strong, local appearance-based prediction.

Bayesian Formulation

\[
\begin{align*}
\text{Posterior} & \propto \text{Likelihood} \times \text{Motion Model} \times \text{Prior} \\
& = p(x_t | z_{t-1}) \times p(x_t | x_{t-1}) \times p(x_{t-1} | z_{t-1}) \times p(x_{t-1} | \theta_{t-1})
\end{align*}
\]

Ground Truth Results

Comparison is made with three existing motion models:

- **Uniform:** A uniform prior over a 61 \times 61 pixel search window.
- **Acceleration:** A constant acceleration model, re-estimated every frame using the last three observations.
- **Median:** The median two-frame motion within a 30 pixel radius.

For each frame, compute:

\[
\int p(x_t | x_{t-1}) p(x_{t-1} | z_{t-1}) \text{d}x_{t-1}
\]

Guided tracking: \( x_t \) from \( z_{t-1} \)

\[
p(x_t | z_{t-1}) = N(x_t | \mu_{x_{t-1}}, \Sigma_{x_{t-1}}) \text{Normal}
\]

For each frame, compute:

\[
\int p(x_t | x_{t-1}) p(x_{t-1} | z_{t-1}) \text{d}x_{t-1}
\]

\[
\text{Diffused Prior}
\]

Then:

- The **diffused prior** is robustified, via blend with unity.
- The **likelihood** is computed as the non-max suppressed NSSD response surface using the patch from the previous frame (for efficiency).
- \( x_t \) is the KLT update from the mode of the posterior.
- \( \Sigma_t \) is set to approximate the posterior by a Gaussian.
- The **diffused prior** is actually a weighted mixture of Gaussians using \( M = \{6, 7, 8, 9, 10\} \).

For all our experiments, \( \gamma = 10 \& R = 6 \).

Data Sets

- Giraffe: 720 \times 576, 100 frames
- Leopard: 540 \times 360, 242 frames
- Mouth: 256 \times 256, 141 frames
- Zebras: 720 \times 576, 171 frames

Algorithm

1st Pass (offline)

- **Fitting 10-frame motion models**

For each frame:

1. Detect interest points.
2. Match interest points in the previous \( M-1 \) frames to generate a measurement matrix, \( B \), containing \( M \)-frame tracks (arranged column-wise).
3. Robustly fit a rank \( R \) non-rigid motion model: \( B \)

That is, we approximate (matrix sizes in brackets):

\[
M(2M \times N) \approx B(2M \times R) \circ (R \times N)
\]

4. Compute

\[
B = B_0 B_0^\top
\]

Note: for brevity, we define \( v = 1 \cdot M-1, \) hence \( x_{1, \ldots, v} \equiv x_{t-1} \).

2nd Pass

- **Proposed vs. Uniform Acceleration Median**

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Proposed</th>
<th>Uniform</th>
<th>Acceleration</th>
<th>Median</th>
</tr>
</thead>
<tbody>
<tr>
<td>Giraffe</td>
<td>45.4%</td>
<td>14.1%</td>
<td>18.1%</td>
<td>6.0%</td>
</tr>
<tr>
<td>Leopard</td>
<td>14.6%</td>
<td>20.0%</td>
<td>11.9%</td>
<td>6.0%</td>
</tr>
<tr>
<td>Mouth</td>
<td>12.6%</td>
<td>3.8%</td>
<td>3.0%</td>
<td>3.8%</td>
</tr>
<tr>
<td>Zebras</td>
<td>12.6%</td>
<td>3.8%</td>
<td>3.0%</td>
<td>3.8%</td>
</tr>
</tbody>
</table>

Track Length. Average improvement of correct track length, in comparison to the three existing motion models, for the four test sequences.

Predictive Power. The average RMS error, in pixels, of predictions using the ground truth data. Percentage improvement achieved by our model is given in brackets. Here, ‘Uniform’ is the constant position model.

Aeron M. Buchanan
Dept. Engineering, University of Oxford, UK
http://www.robots.ox.ac.uk/~awb

Andrew W. Fitzgibbon
Microsoft Research, Cambridge, UK
http://research.microsoft.com/~awf