Real-time human body tracking

Fabrice Caillette

Toby Howard
Goal and applications

- Recover human body pose parameters from camera views only.
- Have the tracking system fast enough for real-time applications on a single machine.
- Robustness to noise, occlusions.
- Applications: Motion capture, human computer interactions, virtual conferencing...
Chosen approach

- 3-D reconstruction
  - Coherent and robust basis for tracking
  - Overhead to minimize
- Body parts recognition with “blobs”
  - Fast and flexible approach
  - Needs a model and correct prediction to be robust
- Motion prediction using both high-level knowledge and model dynamics.
3-D Reconstruction: Visual Hull
3-D Reconstruction: Voxel Projection
Image Segmentation:
Euclidian distance
Image Segmentation: Adding Variance
Image Segmentation: Mahalanobis Distance

\[
\begin{pmatrix} Y & U & V \end{pmatrix} \cdot \begin{pmatrix} \text{Var}_Y & \text{Cov}_{YU} & \text{Cov}_{YV} \\ \text{Cov}_{YU} & \text{Var}_U & \text{Cov}_{UV} \\ \text{Cov}_{YV} & \text{Cov}_{UV} & \text{Var}_V \end{pmatrix}^{-1} \cdot \begin{pmatrix} Y \\ U \\ V \end{pmatrix}
\]

- Full Covariance Matrix for each pixel.
- Slow, especially if we continuously update the model of the background.
- Still some errors.
Image Segmentation: final words

- Robust techniques are usually very slow.
  - Closing-kernel filters to remove noise.
  - Detection and removal of shadows.
  - Adaptive mixture of gaussians...

- Our goal was to achieve real-time performance on a single PC.

- Knowing approximately the pose of the subject can help the segmentation.
Back to 3-D Reconstruction

\[ \text{Mean} = \ldots \]

\[ \text{Variance} = \frac{1}{N} \sum_{i=0}^{N} (\text{distance}_i - \text{Mean})^2 \]
3-D Reconstruction: High Variance

If, in at least one of the views, the variance is higher than an adaptive threshold:
3-D Reconstruction: Low Variance

- If the mean distance to the background is sufficiently low in at least one view:

- If the mean distance to the background is sufficiently high in all the views:

- Otherwise:
3-D Reconstruction: Dealing with unknowns

- After counting the neighbour voxel classes...

  - the unknown voxel takes the class of the majority
3-D Reconstruction: The last cubes...

Regroupment
3-D Reconstruction: Results
Blobs

- A blob is a set of voxels, coherents in their attributes.

- Contains:
  - Mean position and colour
  - Covariance matrices for position and colour
  - Some constraints on size, shape and colour change.
Expectation-Maximization (1)

- The blobs have the parameters from the last frame combined with motion-prediction.

- Distance voxel-blob: Mahalanobis.

\[
\begin{pmatrix} x & y & z & y & u & v \end{pmatrix} \cdot C^{-1} \cdot \begin{pmatrix} x & y & z & y & u & v \end{pmatrix}^T
\]
Expectation-Maximization (2)

- Each voxel is assigned to the nearest blob.

- Blobs have a bounded volume, so that only the nearest voxels are affected.

\[ i = \arg\min_i \left( \text{Distance}(\text{CurrentVoxel}, \text{Blob}_i) \right) \]
Expectation-Maximization (3)

- The blob parameters are re-estimated from their set of voxels.

- If error too big, re-iterate from step (1) with the new parameters.
Model

- Kinematic model with joint constraints, driven by Inverse Kinematics (CCD)

- Blobs created, deleted and placed on the bone-segments dynamically.
A quick word on motion-prediction

- Needed, especially if the framerate is low.
- Linear filters are generally not appropriate because human motion is highly non-linear.
- Particle filters would fail due to the high dimension of the parameter space.
- HMMs seem to have some success.
Motion prediction: what's next?

- Reduction of the parameter space (PCA).
- Model dynamics (Jacobian + Constraints).
- High level knowledge (learning).
- Integration using Kalman filtering.
System overview


Blobs Re-Generation → Model Params. → Motion Prediction

Model Params.
Implementation
Thank you.

Questions?