Person detection, tracking and human body analysis in multi-camera scenarios

Montse Pardàs (UPC)

ACV, Bilkent University, MTA-SZTAKI, Technion-ML, University of Surrey, University of Amsterdam, UPC
Summary

- Multi-camara scenarios definition and applications.
- Multi-level foreground segmentation and tracking
- 3D person tracking with particle filters
- Human motion capture
  - Marker-based
  - Hierarchical annealing
  - Latent-space based
- Examples of other works in the e-team:
  - Moving object detection and classification
  - Eye tracking
Input Data Generation

Real World

Original images

F/B Segmentation

3D (SfS)

3D coloring
Applications

- Multi-person tracking
- Activity recognition
- Head orientation
- Focus of attention
- Human motion capture
- Gesture/Gait recognition
- Raw features
- High semantic level

3D (SfS) → 3D coloring
Foreground detection: multi-level fg segmentation

- Based on a statistical modeling of the pixels’ value \( X_t \) in the \((i,j)\) coordinates, using Gaussian models.

\[
P(X_t) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(X_t-\mu)^2}{2\sigma^2}}
\]

A pixel will be classified as belonging to a Gaussian model if \( |X_t - \mu| < k\sigma \).

- Background model
- Moving foreground model:
  - A pixel not classified as bg or static fg is classified as moving fg.
- Static foreground model:
  - When a moving foreground model has been observed during a certain time \( T \), the model is transferred to the static foreground.

- An incoming pixel can:
  - Match the current foreground model: it is updated and its counter is increased
  - Match the background or static foreground models: it is updated and the counter of the foreground model is decreased
  - Not match any existing model: The foreground model and its counter are re-initialized.
Multi-level foreground segmentation
Mean-shift tracking with multi-level foreground

- Double register, for static and moving objects: centroid position, size, color histogram and counter of appearance

**Foreground objects detection**
- Filtering of the multi-level foreground detection to generate Connected Components if minsize.

**Temporal association of static objects**
- Check if there is a static CC in the position of a static object.
  - Update the corresponding static object register
- Otherwise:
  - The object has re-started its motion if in the area of the static object there is at least 70% background. The static object is transferred to the moving objects register.
  - The static object has been occluded: a moving CC is detected in the area of the static object. No action is taken.
Mean-shift tracking with multi-level foreground

- **Mean-shift tracking of moving foreground objects**
  - The estimation of the position of the objects in the moving object register is performed using mean-shift.
  - The foreground mask is applied to the original image before performing the mean shift.
  - All the CC in an area of the size of the object around the mean shift estimation are assigned to the object. The object features are updated.
  - If two or more moving objects share the same CC, we enter an occlusion situation and the update is done only on the centroid value using the mean shift estimation.

- **Detection of new Static and Moving objects**
  - The CC which have not been associated to any object are introduced as new objects.
Mean shift tracking with multi level foreground
Person tracking in multi camera scenarios

- Perform multi-person tracking in an indoor scenario employing 3D information (SfS+color)
- Particle Filters based solution
- Evaluate the performance of this method by using a standardized database (CLEAR 2007)

**Particle filters:**
Key idea: represent the posterior \( p(x_k|z_{1:k}) \) by a set of random samples.
Four steps are followed: resampling, propagation, evaluation and estimation
When implementing a PF, two issues are to be taken into account:
Likelihood evaluation: how to assign weights to our particles
Propagation model: how to “move” our particles to efficiently sample the state space
Likelihood evaluation:
- Every particle defines a possible location of the person represented by an ellipsoid
- For every person a reference color histogram is built up
- Weight assigned to a particle is a function of its overlap with the binary 3D reconstruction and its color similarity with the reference histogram (Battacharyia distance)
Multi-Person Issue

- Multiple targets may be tracked using a PF with a state containing the 3D position of each of them. However, complexity makes it unfeasible.
- A PF is assigned to every target and an interaction model is defined.
Results
Results

Accuracy

Best performing tracking algorithm at CLEAR 2007 Evaluation Campaign

Precision
Multimodal integration

- Audio information was added to produce a more robust tracker
- However, only information fusion at feature level has been conducted so far, giving slightly better results (5%)
- Information fusion at data level is under study
Two approaches are widely employed: marker and makerless.

- It provides information related with the pose of the person. A human body model is assumed.
- It poses a technology challenge due to the high dimensionality of the state space (~22 DOF). Moreover, this space is highly non-convex.
Marker based capture

- Intrusive but provides very precise results
- Widely used by the cinema industry. Quite expensive equipment (>10K$) and requires dedicated hardware
- An annealing PF together with off-the-shelf hardware solution has been developed
Markerless based capture (I)

- Non intrusive but more challenge since there are no artificial aid (markers)
- Still an open problem for unconstrained motion
- PF definition: every particle represents a possible pose of the person
- Likelihood evaluation: defining a fitness function matching the input data against the particle pose taking into account 3D information and color
Markerless based capture (II)

Original images

Voxel reconstruction

Particle $x_t^m - \beta(x_t^m)$

$\beta(x_t^m) \cup z_t = 0.84$

Particle $x_t^n - \beta(x_t^n)$

$\beta(x_t^n) \cup z_t = 0.62$
Markerless based capture (III)

- Structural annealing PF technique that performs a layered fitting of a hierarchical model to the input data.

- Input data is very corrupted: a simple body model is used.
- Data is more complete: the full articulated model is used.
- Legs are not visible, hence they are removed from the analysis model.
Markerless based capture (IV)
The drawback in the annealing particle filter tracker is that a high dimensionality of the state space causes an exponential increase in the number of particles that is needed to be generated in order to preserve the same density of particles.

We use a set of poses in order to create a latent space with a low dimensionality. The poses are taken from different sequences:

- Walking
- Running
- Punching
- Kicking

We perform non-linear dimensionality reduction using Gaussian Process Dynamic Model (GPDM) and construct a latent space.
The Tracking Scheme

Latent to Model Mapping

Model to Camera Projection

Technion-Israel Institute of Technology
Multi-Action Learning

2D latent space from 3 different motions: lifting an object (red), kicking with the left (green) and the right (magenta) legs.

Technion-Israel Institute of Technology
Results
Other works:

Moving object detection and classification
Other works:

- Eye tracking