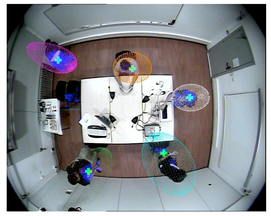
# Particle Filtering and Sparse Sampling for Multi-Person 3D Tracking

C. Canton-Ferrer, R. Sblendido, J.R. Casas, and M. Pardàs
Technical University of Catalonia, Barcelona - Spain

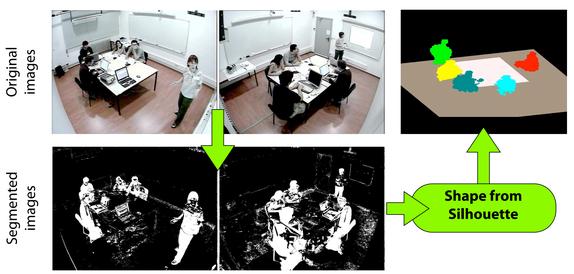
## 1. Introduction



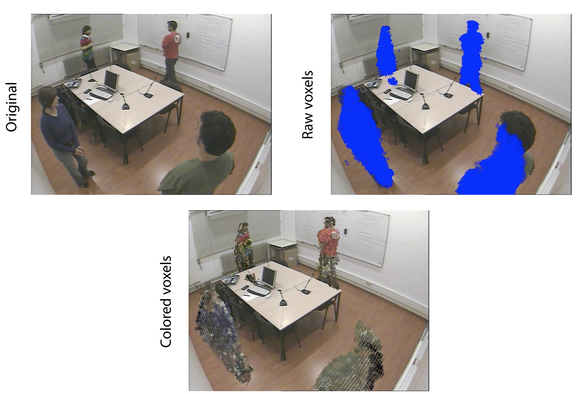
**Where and Who?**

**Where is each person in the scene? Who is who along time?**

- **GOAL:** Tracking a number of people and keeping a consistent track of their identities along time in a SmartRoom scenario equipped with multiple cameras.
- Real time performance is desired hence fast algorithms are required.
- Redundancy among views is exploited to generate a 3D reconstruction of the scene by means of a voxel representation.




- Images are segmented using Stauffer's background substraction algorithm and silhouette consistency among cameras was used to assess voxel's occupancy.
- Color information is added to the voxel reconstruction.




## 2. Particle Filtering

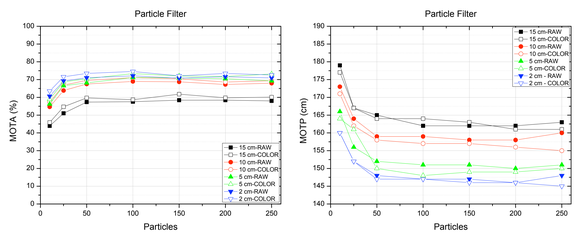
- A Particle Filtering (PF) strategy is devised to track a number of targets in the scene using the 3D colored voxel reconstruction.
- The likelihood of a particle describing an instance of the human body is the only considered defining parameter of the PF.

**Likelihood evaluation**

- An adaptive reference histogram of each target in CbCr space is available. An ellipsoid describing the human body is associated to a given particle.
- Two factors are linearly combined in the likelihood evaluation:
  1. **Overlap:** How much of this ellipsoid overlaps with the input data?
  2. **Color:** How much do the colors of the data voxels enclosed in this ellipsoid match the reference histogram of this target?

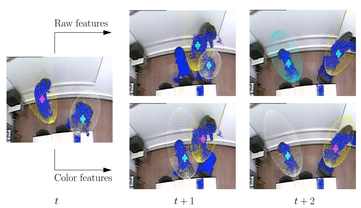
**Evaluating this likelihood is computationally expensive!**
**Cost: $O(N_{particles} \cdot \text{Volume Ellipsoid})$**
**Real-time might be unreachable!**

- **Performance:** The PF algorithm has been tested with the CLEAR 2007 Dataset. The analyzed scenario is a SmartRoom with an average of 5 moving people inside. 5 calibrated and synchronized cameras at 25 fps with a resolution of 768x576 pixels were used. 3 hours of data were analyzed. Two metrics are employed: the **MOTA**, that accounts for the accuracy of the tracker (the percentage of time where you track correctly all targets in the scene) and the **MOTP**, that scores the precision of your centroid estimation of all targets in the *xy* plane.




- MOTP score has a noticeable dependency with the employed number of particles and the size of the voxel. Color information improves the results in comparison with the usage of raw information.

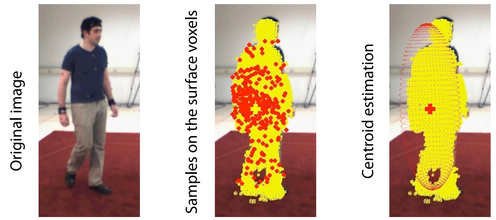
## 3. Color information usage




Color information allows (in both algorithms) disambiguating the identity of targets after a cross-over.

## 4. Sparse Sampling

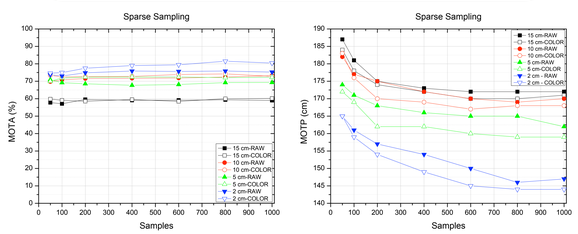
- Sparse Sampling (SS) technique is an alternative to PF for 3D person tracking with a lower computational cost. A set of samples placed on the surface of the person blob allows estimating the centroid of the target. We are no longer sampling a state variable (centroid) as done in the PF.
- The position of samples evolve with time following the PF steps: resampling, propagation, evaluation and estimation.




**Sample Likelihood Evaluation**

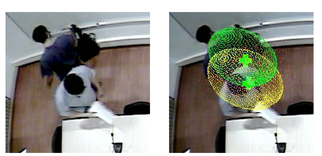
- Is computed in a similar way to PF. Likelihood is computed in a spatial neighborhood around the sample position and has two contributions:
  1. **Surface:** Samples attain its maximum value when half of its neighbors are empty and the other half occupied.
  2. **Color:** A local histogram is matched against the reference histogram.

**Evaluating likelihoods over a neighborhood is cheaper!**
**Cost: $O(N_{samples} \cdot \text{Size Neighborhood})$**
**Real-time might be acchieved!**




- **Performance:** Due to the discrete resampling, samples adapt better to noisy data thus reaching higher MOTA scores while MOTP has similar results than PF.

## 5. Tracker interaction



A particle/sample blocking method is proposed to drive the interaction of particles/samples when some targets get close to each other.

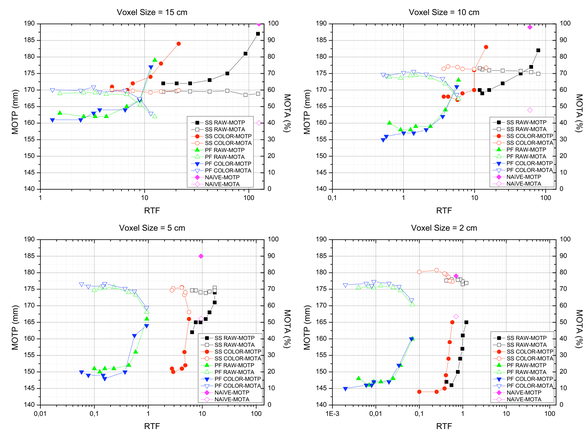
## 6. Comparison

| Method | MOTA (%) | MOTP (mm) |
|---|---|---|
| Face detection + Kalman filtering (Katsarakis & Talantzis, 2007) | 59.66 | 91 |
| Appearance model + PF (Lanz & Chippendale, 2007) | 59.62 | 141 |
| Upper body detection + PF (Bernardin et al., 2007) | 69.58 | 155 |
| Zenital camera analysis + PF (Bernardin et al., 2007) | 54.94 | 222 |
| Voxel analysis + Heuristic tracker (Canton et al, 2007) | 30.49 | 168 |
| **Voxel analysis + PF** | **74.56** | **147** |
| **Voxel analysis + SS** | **81.5** | **144** |

Comparison with other systems evaluated using the same dataset and the same metrics.

## 7. Real-Time Performance

- Tracking multiple targets with a real-time performance is a desirable property for a tracking system
- **Performance:** Computational load of the algorithms increase as the voxel size decreases. SS is, in all cases, faster than the PF one. Higher values of RTF are desirable.




## 8. Conclusions

- Two algorithms are presented for the 3D multiple person tracking task using the information gathered by multiple cameras. A voxel reconstruction approach is followed.
- The PF algorithm might not achieve real-time performance due to the involved complexity when evaluating the likelihood.
- The SS algorithm is presented as an alternative reaching real-time computation and a higher performance than PF.
- Color information allowed resolving mismatches among targets.