1. Introduction
- **Scenario**: multiple rigid objects moving on a ground plane tracked by a moving camera (rotation + translation)
- **Goal**: infer 3D volumetric shape of both moving objects and static background across multiple frames
- **Applications**: video surveillance, image synthesis, etc.

2. Initial processes of the approach

3. Volumetric decomposition
- Adaptive volume tessellation
- Hierarchical octree
- Moving objects contain more smaller voxels
- Dynamic voxel labels

4. Photo-motion variance measure
- Evaluate the consistency between assigned voxel labels and the original images
  \[
  \phi(v, t, v', t') = \lambda \phi_{photo} + (1 - \lambda) \phi_{motion}
  \]
- Photo variance: matching multi-oriented image patches
- Motion variance: overlap of voxels’ bounding boxes

5. Combined voxel labeling process
- Surface point initialization
- Multiple trials with different thresholds
- Sequentially “local” decision making
- Energy minimization enforcing global consistency and smoothness

6. Experimental results

7. Summary
- Contributions
  - Photo-motion variance measure
  - Voxel coloring + Graph Cuts
  - Practical multi-stage system
- Future directions:
  - Reconstruction of non-rigid objects (e.g. human)