Joint Segmentation and Tracking of Object Surfaces in Depth Movies along Human/Robot Manipulations

Babette Dellen, Farzad Husain and Carme Torras

Barcelona. Spain

Institut de Robòtica i Informàtica Industrial

23/02/2013

Introduction

- An important area in the field of 3-D vision is segmentation and tracking of depth data.
- Data from the sensors needs to be structured in a way that makes task-relevant visual information more accessible.

Introduction

- A novel framework for joint segmentation and tracking of object surfaces is presented.
- Practical application with low-cost depth sensors.

Recent Work

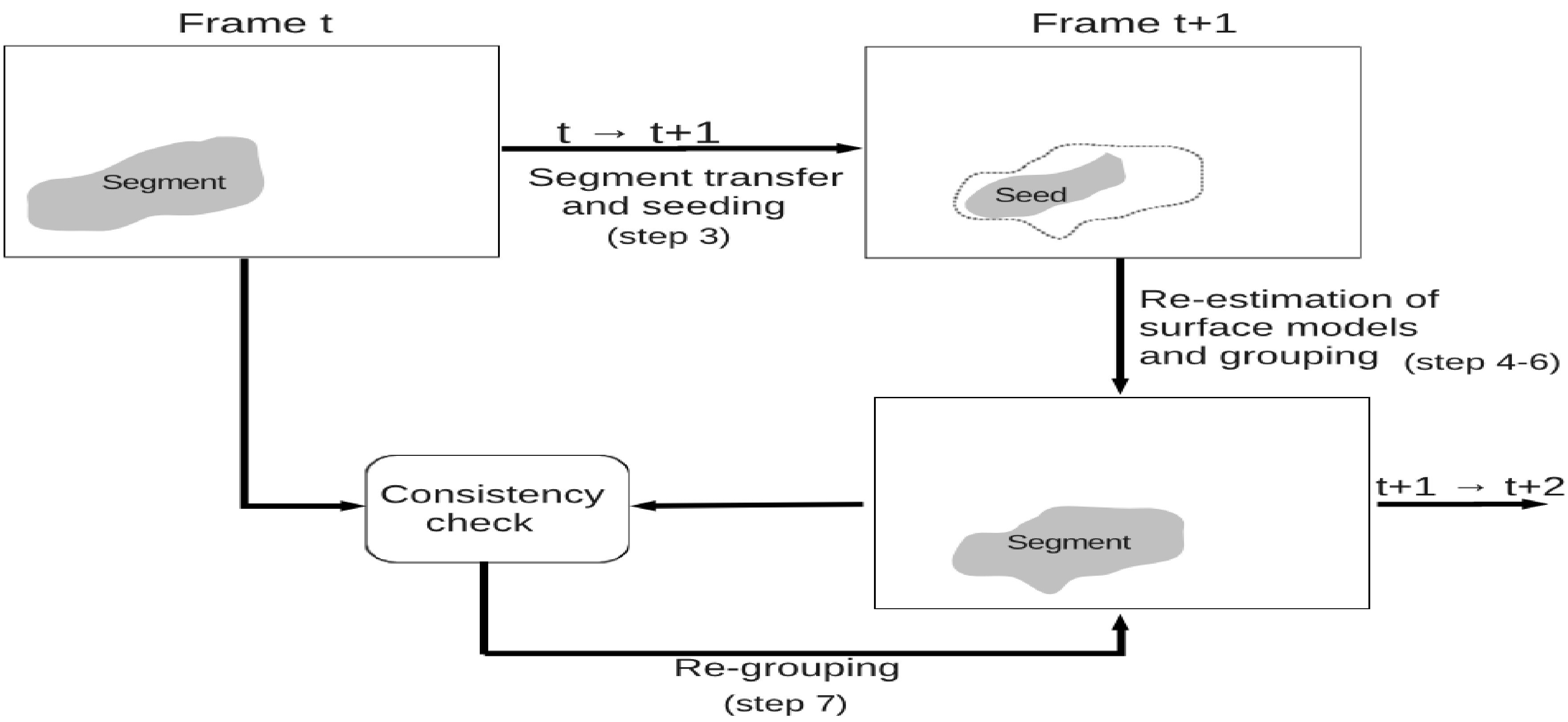
- Depth based segmentation.
  - Using local surface descriptors.
    - Pulli et al., 1993, Hedge et al., 2011, Bab-Hadiashar et al., 2006, Jiang et al., 2000.
- Segmenting and tracking using depth data as a primary cue.
  - Segmenting and tracking particular surface shapes.
    - Pravizi et al., 2008, Ghobadi et al., 2007.
- Primary focus has been on color based segmentation and tracking.
  - Abramov et al., 2010, Deng et al., 2001, Patras et al., 2001, Wang et al. 2009.

Motivation

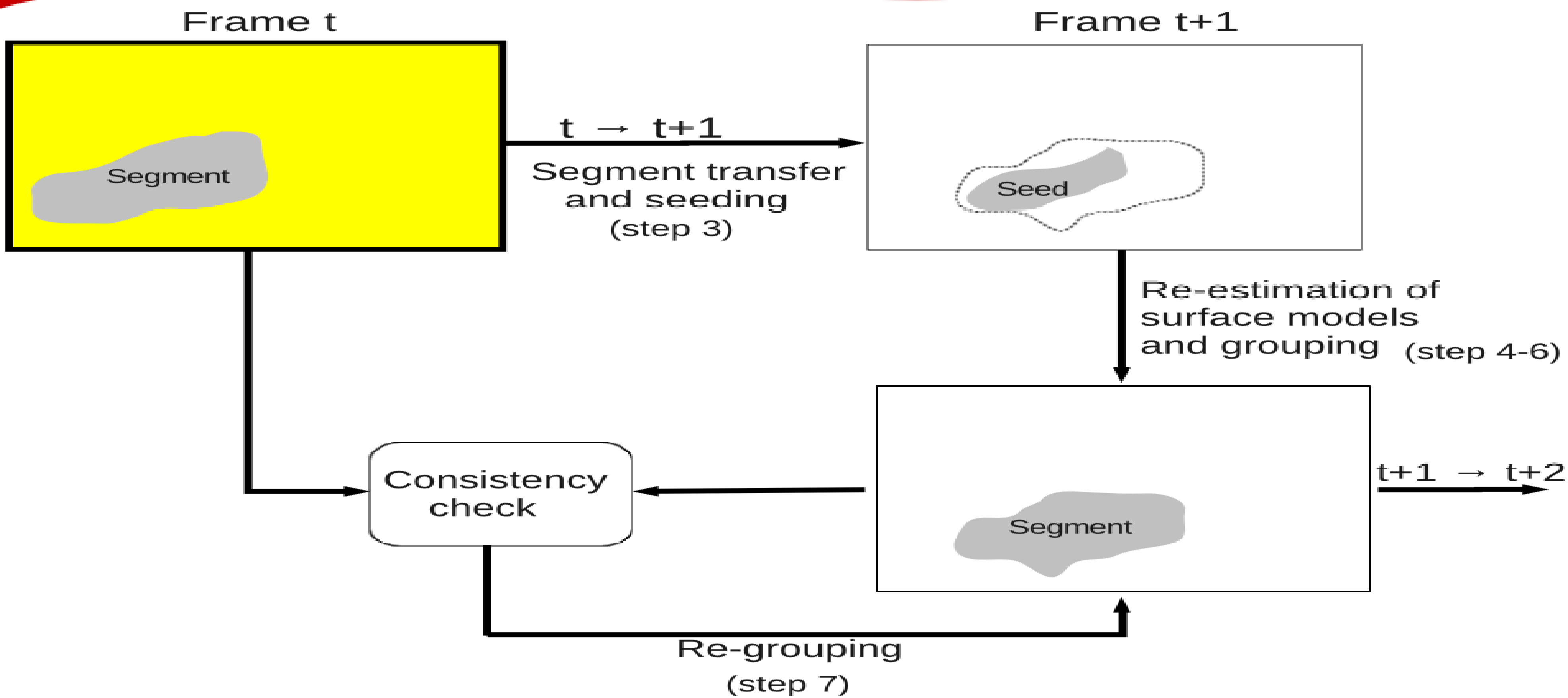
- Existing segmentation algorithms rely on local geometric information such as
  - Surface normals.
    - Jiang et al., 2000.
  - Jump Edges.
    - Han et al., 2004.
- Local geometric properties do not give much information about the location of the surface.
- We determine global surface model parameters, which encode how sampled-points are embedded in 3d-space.

Main idea

- Compute an initial segmentation using color and depth data.
- Transfer previous frame labels to next frame and refine quadratic surface parameters of each segment.



Initial Segmentation



- An initial Segmentation is computed for the first frame.

Initialization

- An initial labeling $l^t(u, v)$ for the first frame is computed using a method, as proposed in [Dellen et al., WACV, 2011].
- A quadratic surface model $f_j^t(x, y)$ is used to fit data corresponding to every segment.

$$z = ax^2 + by^2 + cx + dy + e$$

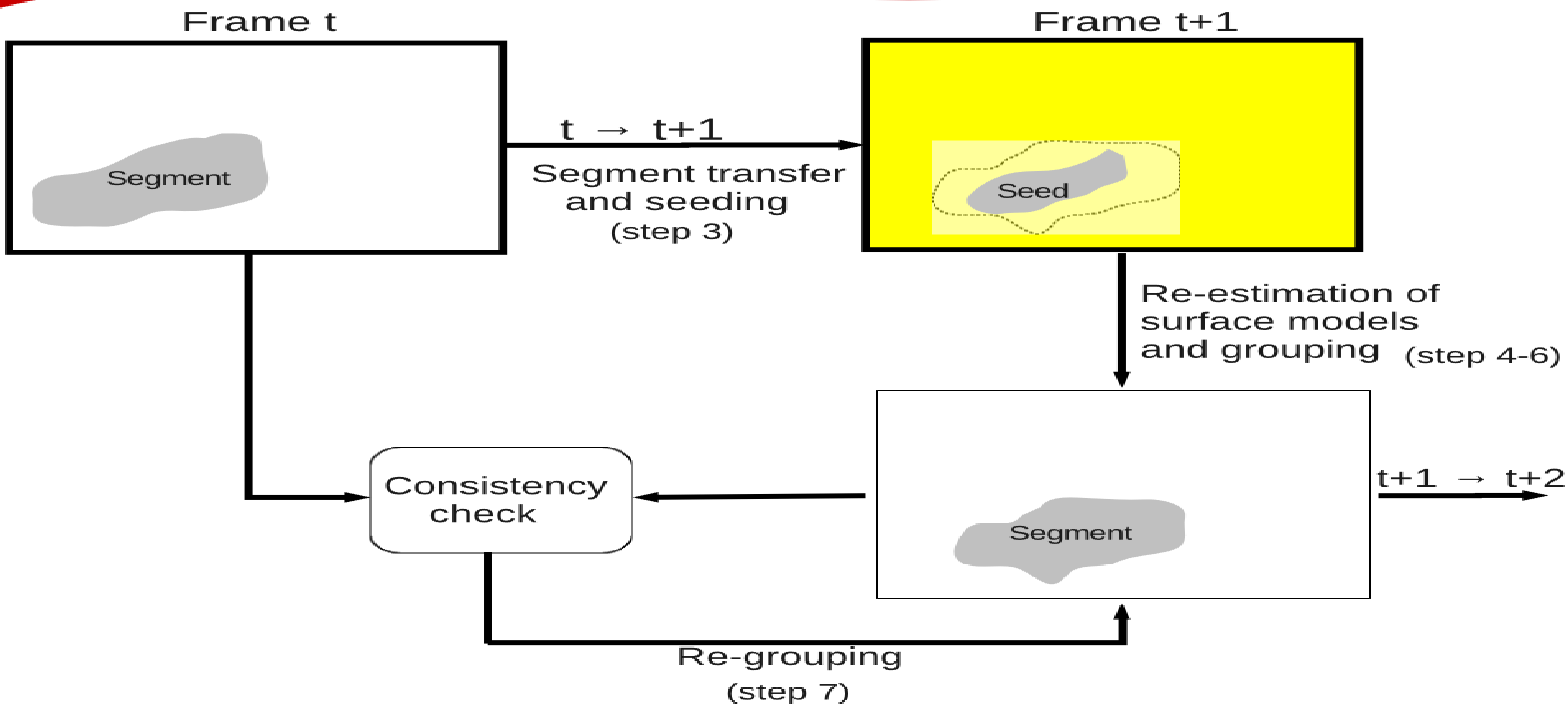
Minimization

- Surface parameters are determined for each segment by performing a Levenberg-Marquardt minimization of the mean square distance.

$$E = 1/n_j \sum_{(u,v)\in s_j} [z_e(u,v) - z(u,v)]^2$$

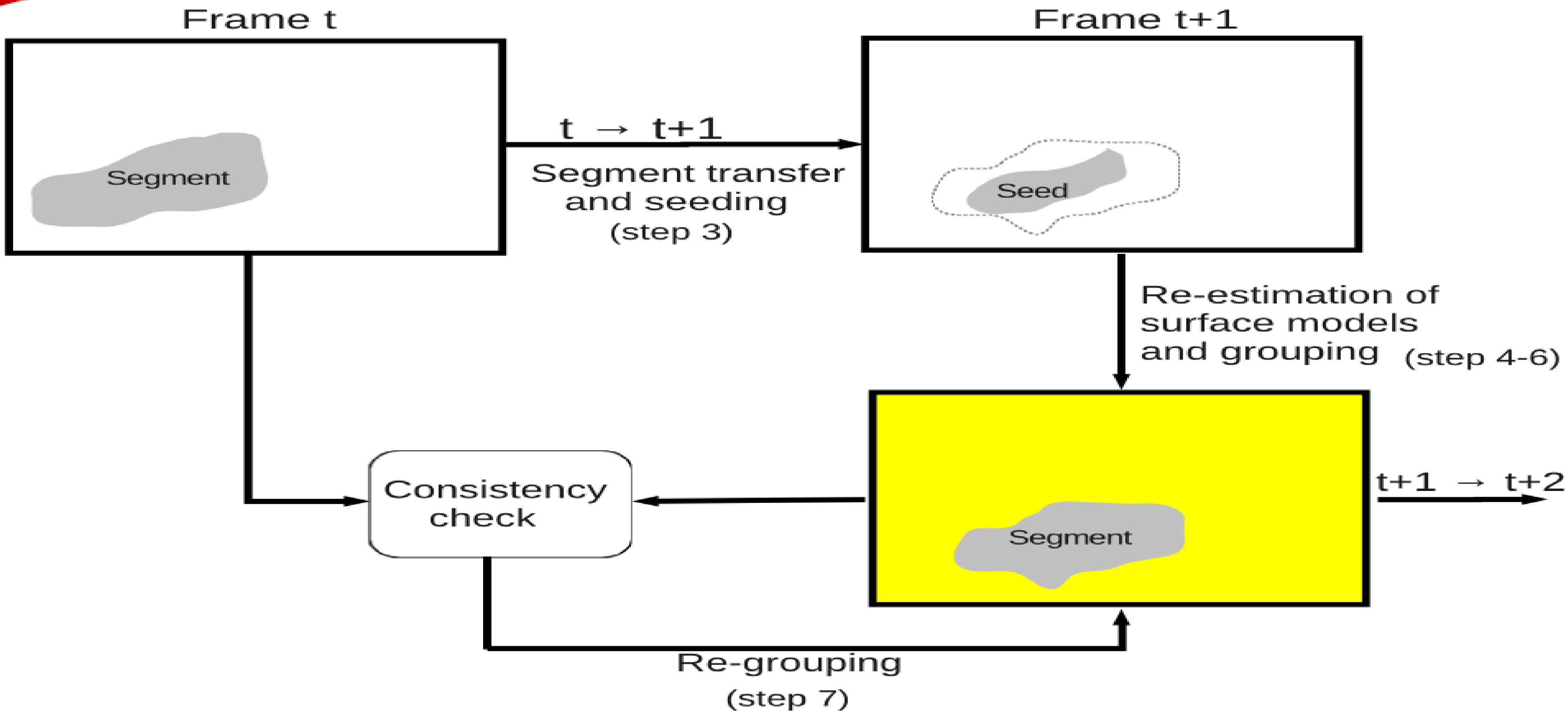
$$z = ax^2 + by^2 + cx + dy + e$$

Seeding



- For each point $p = l^t(u,v)$ of frame t + 1, we find the projected label (u,v).
- Unlabel the points that do not fit the surface (seed generation).
- Update the model parameters by applying the model fitting procedure.

Updating

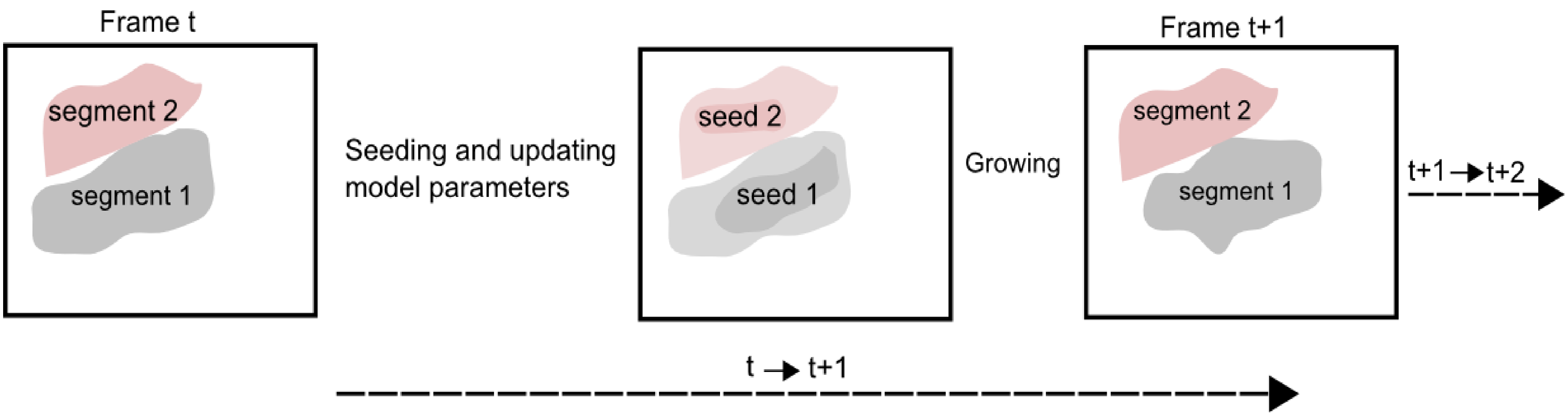


- Relabel the non-seed points based on the updated surface models parameters.

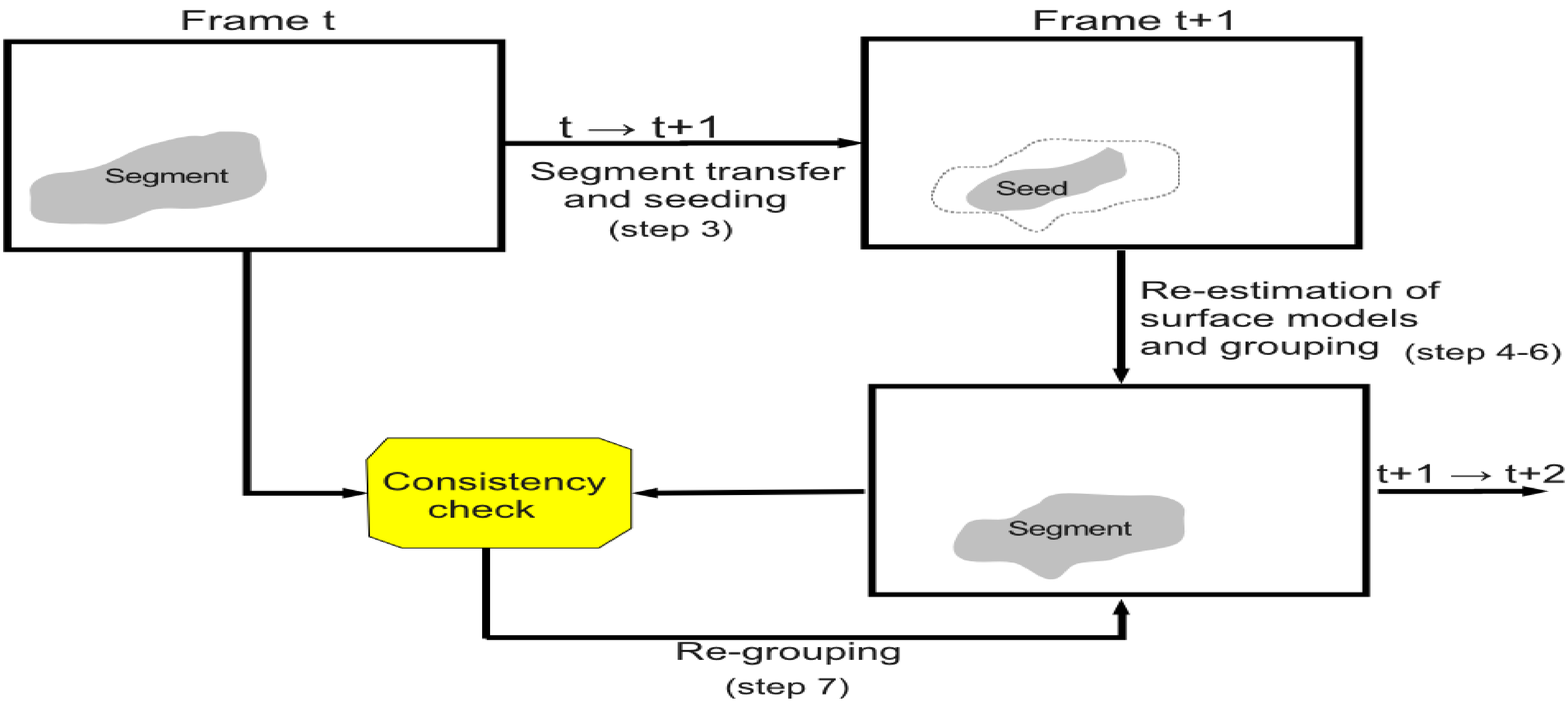
$$d_q(u,v) = |f_q^{t+1}[x(u,v), y(u,v)] - z(u,v)|$$

$$l_j^{t+1}(u,v) \quad = \quad \arg[\min(\{d(l_1), d(l_2), ..\})]$$

Seeding and Relabeling

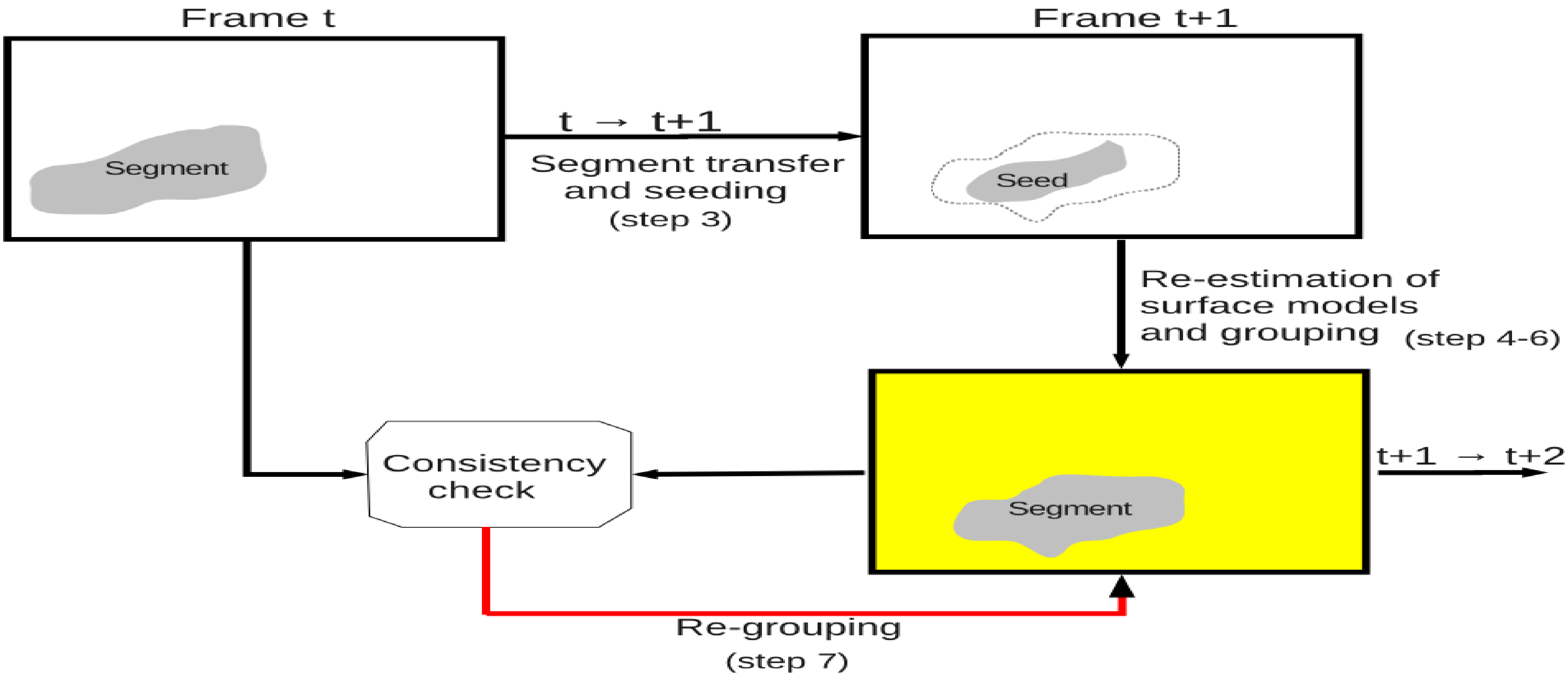


Checking for Consistency



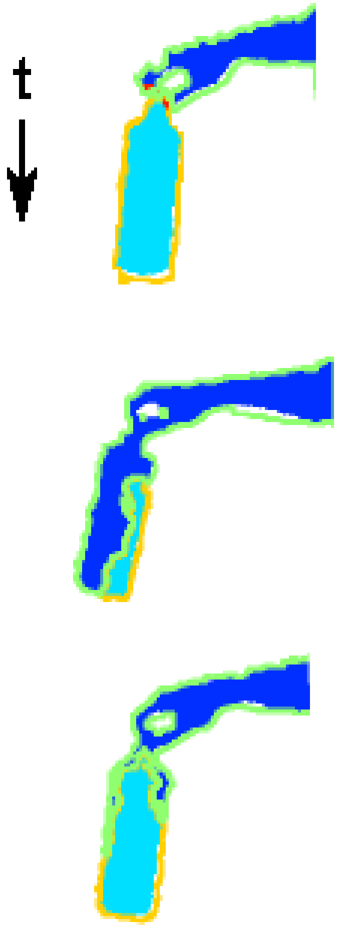
- Assume relatively small motion of objects between consecutive frames.

Regrouping



- Regrouping to maintain temporal consistency.
- Points are re-labeled with one of the segments in their vicinity.

Regrouping



- Segmented hand
- Segmented bottle

- Segments are not allowed to grow or shrink out of proportion

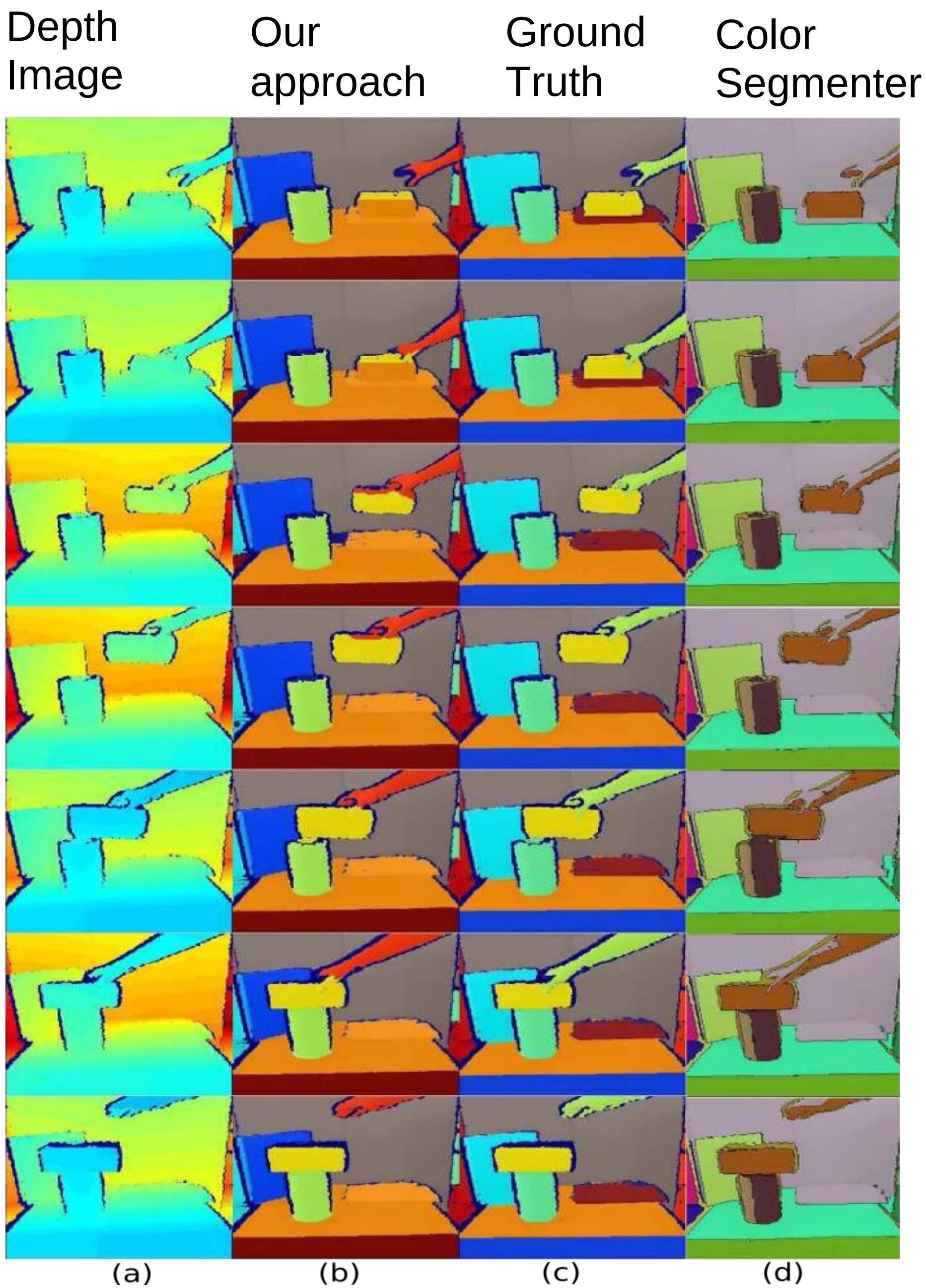
Results

http://www.iri.upc.edu/people/bdellen/Movies.html

Depth Image	Our approach	Ground Truth	Color Segmenter

Results

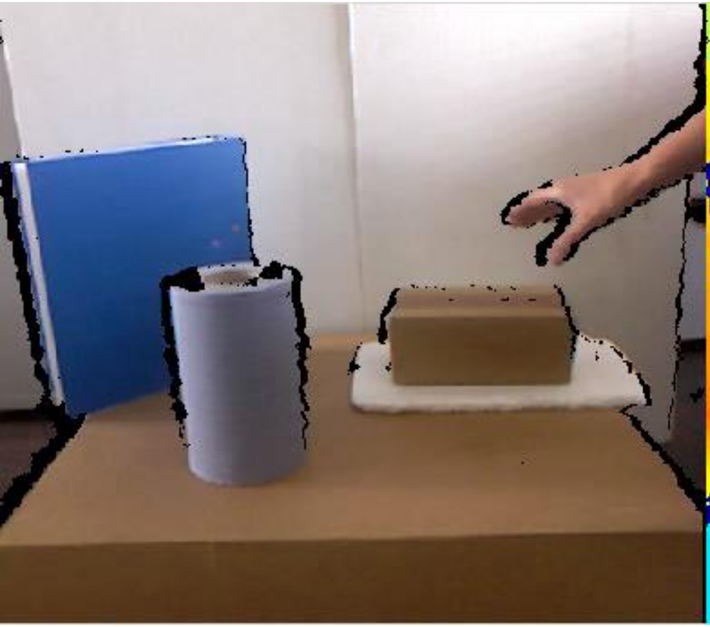
http://www.iri.upc.edu/people/bdellen/Movies.html



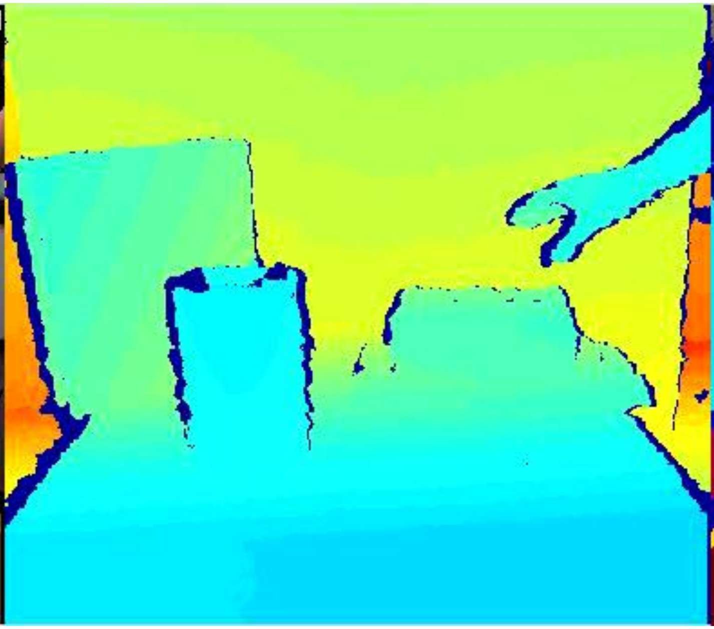
Results

http://www.iri.upc.edu/people/bdellen/Movies.html

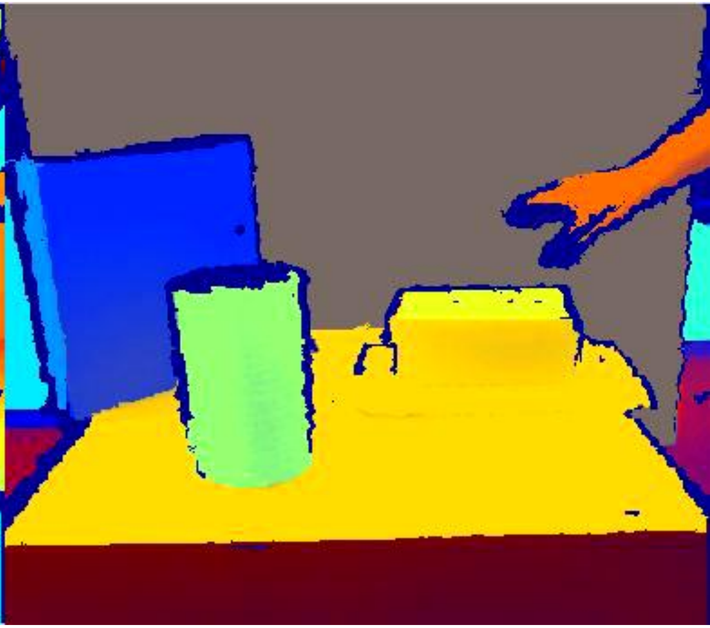
Color Image



Depth Image



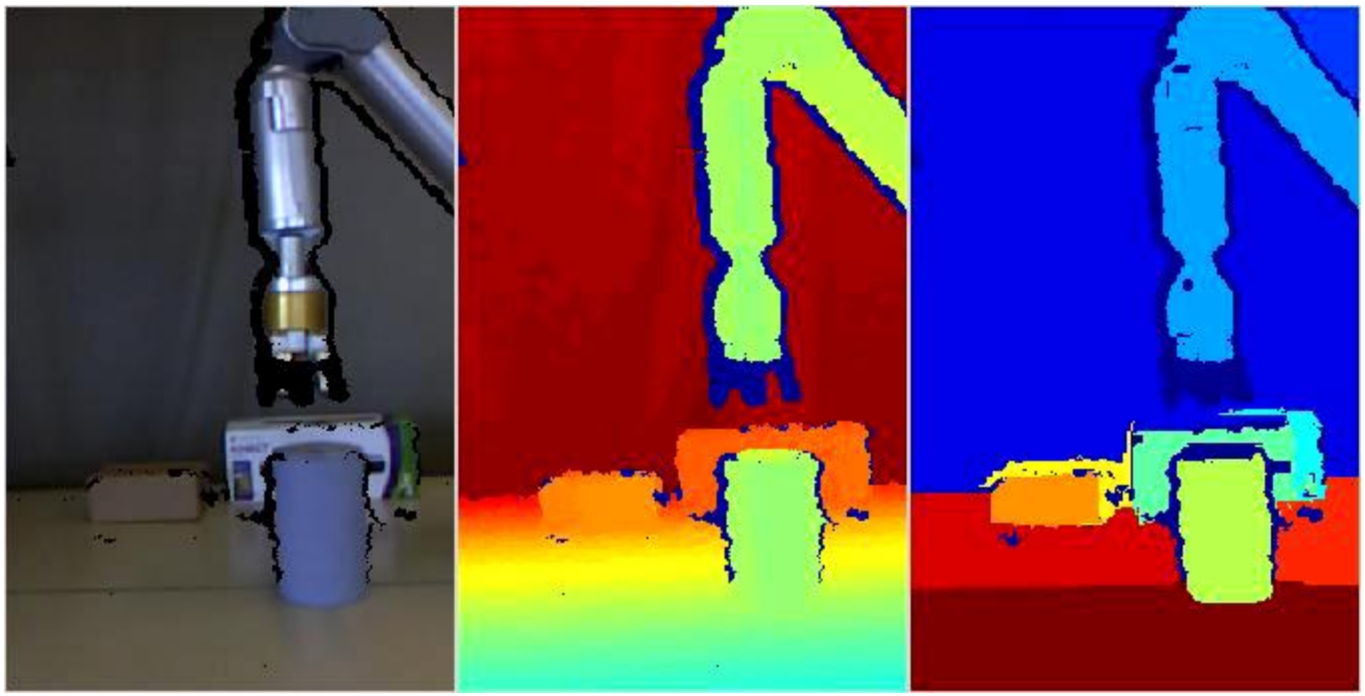
Our approach



Results

http://www.iri.upc.edu/people/bdellen/Movies.html

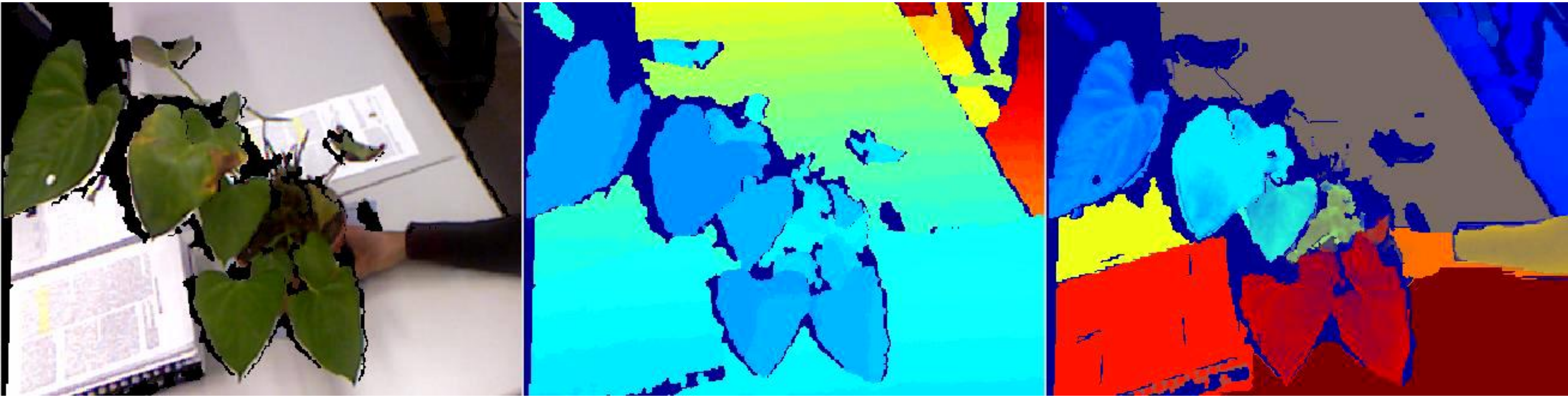
Color Image | Depth Image | Our approach



Results

http://www.iri.upc.edu/people/bdellen/Movies.html

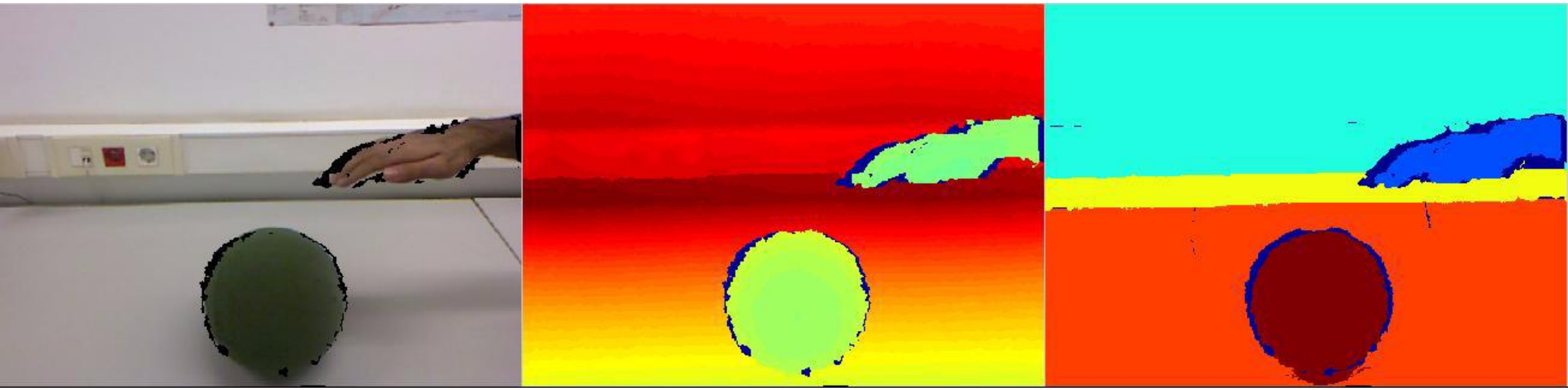
Color Image | Depth Image | Our approach



Results

http://www.iri.upc.edu/people/bdellen/Movies.html

Color Image | Depth Image | Our approach



Performance Evaluation: Segmentation coverage

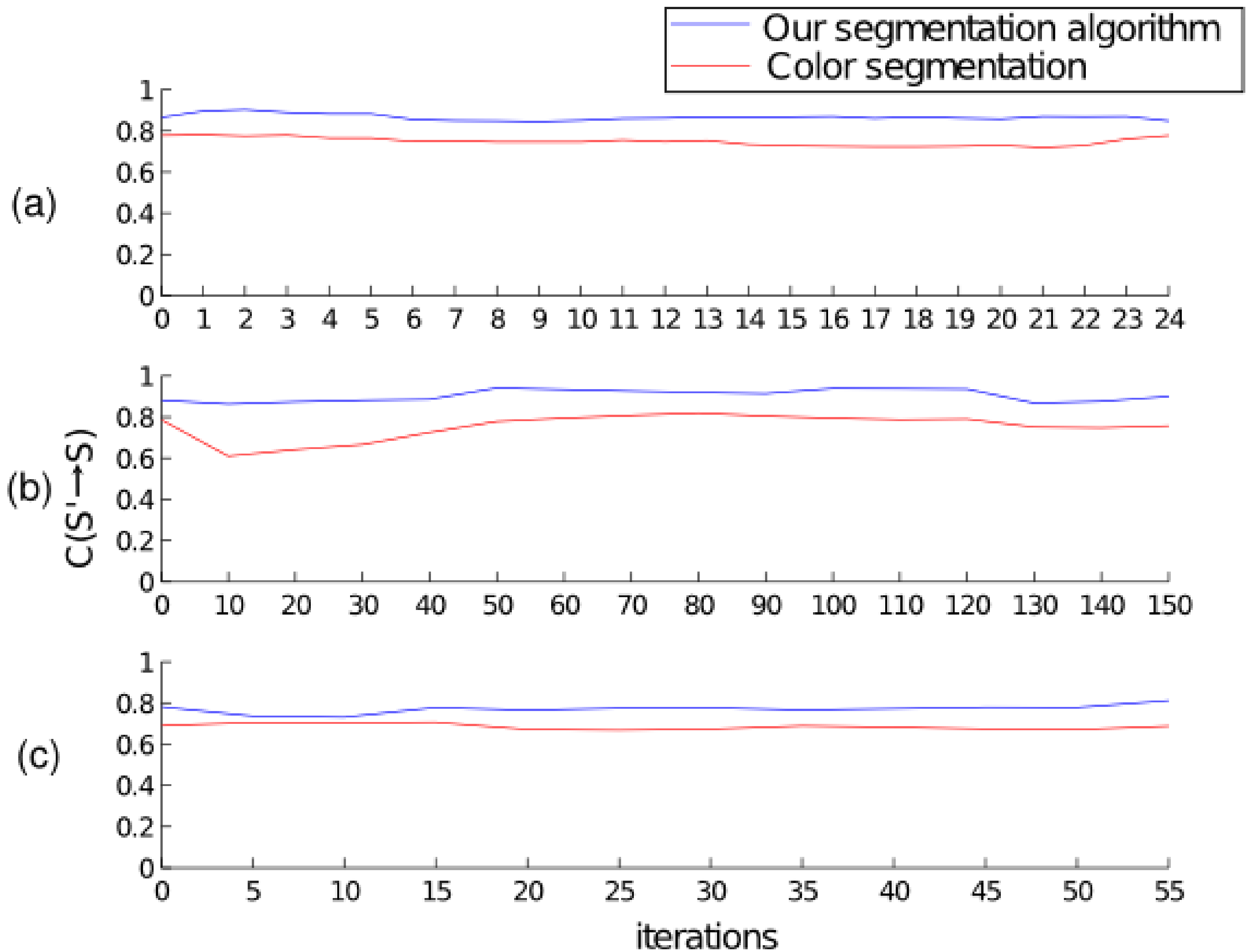
$$C(S' \rightarrow S) = \frac{1}{N} \sum_{R \in S} |R| \cdot \max_{R' \in S'} O\left(R, R'\right)$$

where N is the total number of pixels in the image, $|R|$ the number of pixels in the region R, and $O(R,R')$ is the overlap between the regions R and R' defined as

$$O(R,R') = \frac{|R \cap R'|}{|R \cup R'|}$$

Arbelaez et al. (2009)

Performance Evaluation



-Grundmann et al., 2010

Conclusion

- The algorithm allowed us to segment and track the main object surfaces in the scene.
  - Noise in depth data from Kinect camera.
  - Frequently occurring occlusions.
- Problems that we will address in the future.
  - Depth differences between surfaces are too small, resulting in assignment conflicts that cannot be resolved by the method as it is.
  - Generating new segments in addition to the ones that have been determined in the first frame.

Thank You