3D scene representation in MSEE

Guy Rosman
Sensing, Learning and Inference Group
MIT / CSAIL

April 25, 2014

Scene Modeling in MSEE I

Focus on 2D/3D scene model for predicate processing. I'd like to thank, among others

- Jonathan Balzer - 3D camera and ground plane reconstruction.
- Jason Chang - background color estimation, tracking.
- Randi Cabezas - predicate framework SE on SLI side.
- Many others on the SLI team - Sue Zheng, Chris Dean, Julian Straub, Giorgos Papachristoudis, Zoran Dzunic, Oren Freifeld, John W. Fisher III.

2D vs 3D operators in computer vision I

3D information is useful for many predicates in computer vision. These include, in MSEE:

- Clear line of sight, Occluding
- Below, On, Closer, Farther, Together
- Running, Sitting, Standing, Stopping, Turning, Walking, Crawling, Stationary, Entering, Exiting,

2D vs. 3D operators in computer vision I

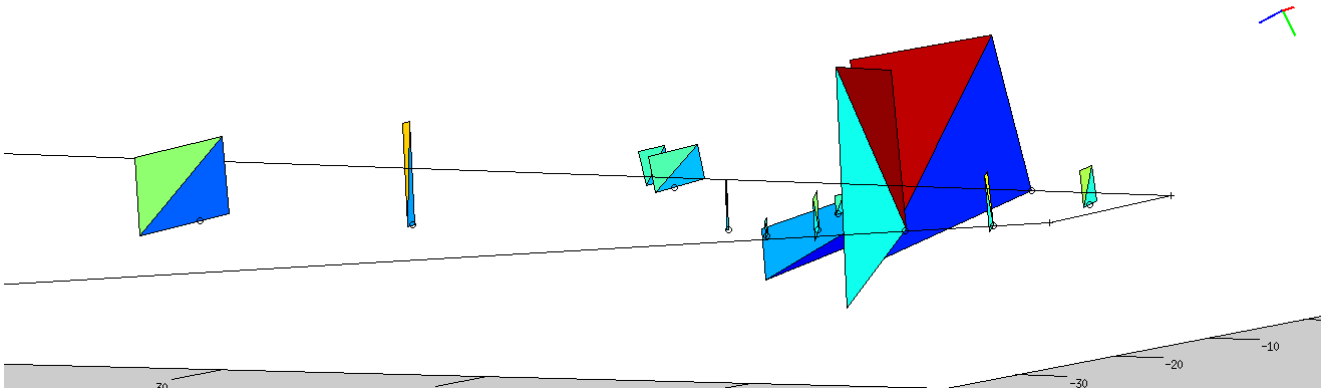
- 2D operators often substitute 3D ones – and depend on priors (classic examples: fronto-parallel assumption in Lucas-Kanade tracking, multi-target tracking).
- This depends on the quality of the reconstruction available, and estimated uncertainty.
- Reasonable treatment of uncertainty exists in SLAM and tracking literature, but not in general computer vision.

Multiview Scene Understanding I

[Bill triggs, 00'] on common errors in bundle adjustment – on the relation between planning the viewpoints selection, and reliability of the model. "Any sequence can be used": Many vision workers seem to be very resistant to the idea that reconstruction problems should be planned in advance, and results checked afterwards to verify their reliability. System builders should at least be aware of the basic techniques for this, even if application constraints make it difficult to use them. The extraordinary extent to which weak geometry and lack of redundancy can mask gross errors is too seldom appreciated..

Generating a scene model

- Simplified scene model - planar ground + fronto-parallel objects
- For full 3D objects, additional priors, or multi-view reasoning are required.
- Triangulation/Multi-view information is not used for object reconstruction.
- The bounding box of each object is used to create a 3D, fronto-parallel flat model of the object.
- Only in scenes where 3D estimation made sense - one place where uncertainty should be quantified.



Building a Scene Model I

Scene model construction – main steps:

1. A background model is constructed for each view.
2. Objects are categorized as planar/vertical (and sky).
3. For each vertical object, a ground point is estimated.
4. A fronto-parallel representation of the objects is obtained based on the camera matrix, ground point location, and bounding box.

Background Model Construction I

Background model estimation needs to cope with several problems

1. Moving background objects - moving leaves.
2. Long-term moving objects.
3. Periodic appearance changes - day / night / cloud cover effects.
4. Other weather effects - rain/hail/snow/fog.

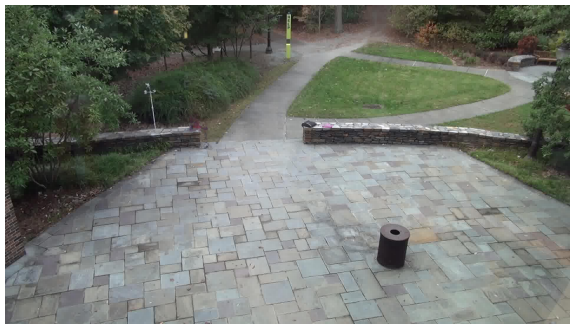
- In our implementation we dealt with the first two effects.
- Item 1 - often addressed in the literature.
- Items (2,3) can be handled with sufficient data.

Background Model Construction I

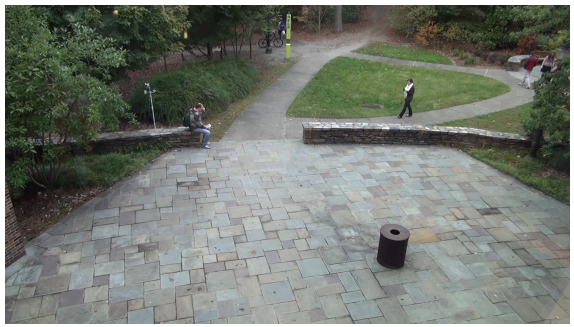
- Preliminary model was obtained by temporal median filtering of the video sequence.
- A more adaptive model was obtained by updating each pixel after acceptance/rejection test, as we track scene movers.
- This creates per frame a background appearance model, which allows us to segment and track moving objects, with less complicated object prior in the tracking.
- Processing is done at each camera separately.
- This is only for static cameras -
  - Handling monocular moving cameras is partially possible.
  - Using multiview can definitely help.

Background Model Construction I

- Temporal median filtering of an hour-long footage



Background model



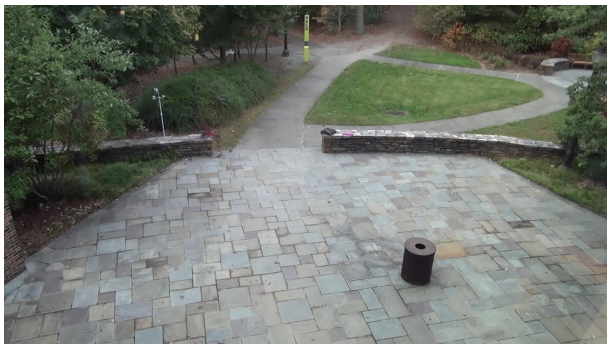
Example image

Log probability

- Some objects remain - long-term moving objects.
- Inaccurate color model due to changes in the intensity - both due to day time /weather, and due to camera gain control.

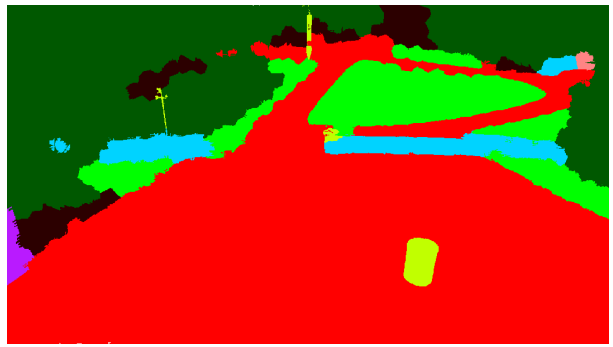
Object model construction I

- Super-pixel segmentation of the background image allows easier (yet manual) annotation of the background into MSEE object classes.
- Connected components used to separate the different objects.
- We assume the bottom of each object touches the ground, and that the bounding box represents the object's silhouette. 
- Process can be significantly automated, perhaps not completely, but is done as part of system setup.



Background image

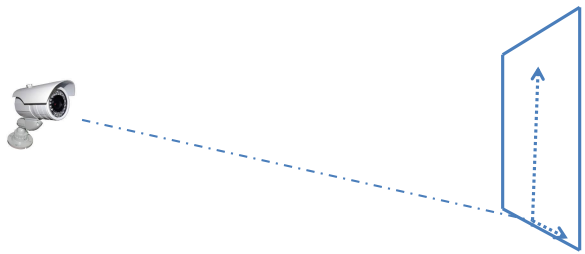
Superpixels



Annotation

Front-parallel Object Model I

- Construct a line l from the camera center to the ground plane intersection point.
- Create plane equation (n, d) with $n = \frac{l}{\|l\|}$ as its normal.
- Intersect the 4 corners of the bounding box with (n, d)
- Construct a planar patch / 2 triangles.



Point-to-Plane Projection I

- If we are given a camera matrix $P \in R^{3\times 4}$, and plane equation (n_g, d_g).
- We mark by $(P)_i$ the rows of P.
- $X \in R^4$ marks the 3D coordinate, in homogeneous coordinates.
- x, y mark the camera pixels.
- We have 3 equations and the assumption that the object is not at infinity.

$$\frac{(P)_1 X}{(P)_3 X} = x, \frac{(P)_2 X}{(P)_3 X} = y, (n,d)X = 0, (X)_4 \neq 0 \rightarrow$$
$$((P)_1 - x(P)_3)X = 0, ((P)_2 - y(P)_3)X = 0, (n_g, d_g)X = 0, (X)_4 \neq 0 \rightarrow$$
$$\tilde{P}_{1..3} X_{1..3} = -\tilde{P}_4$$

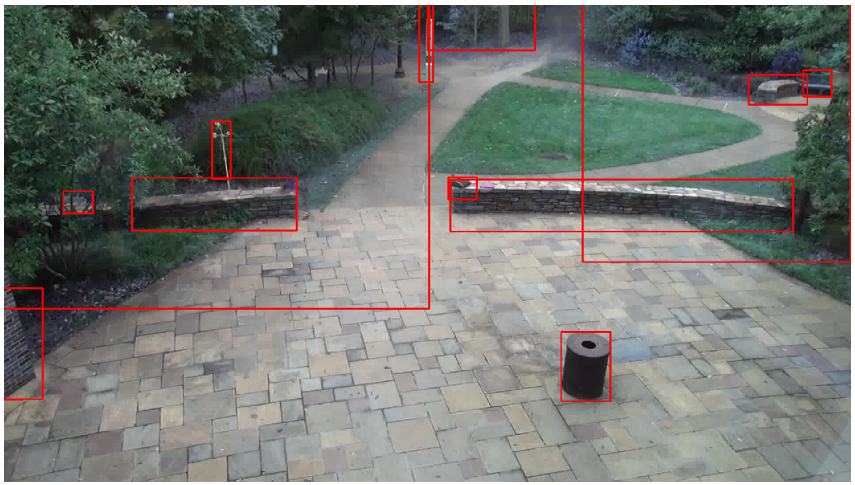
Object model construction I

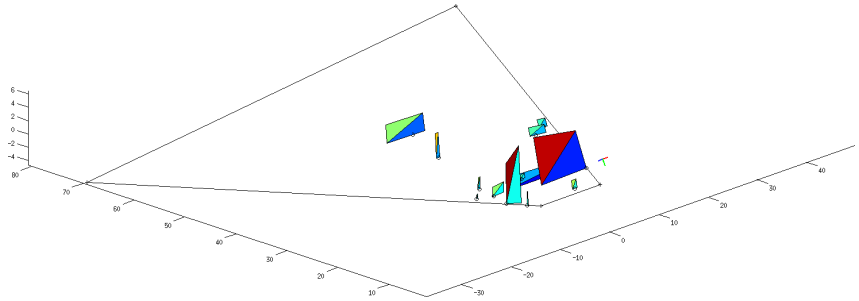
Compared to a single camera + plane assumption, multi-view can

- Correct range estimates where segmentation is faulty.
- Allow a more complete object model with volume, for line of sight reasoning.
- Help reason about and fix background objects connectivity.

Down-sides: require multiple view, sensitive to camera estimation.

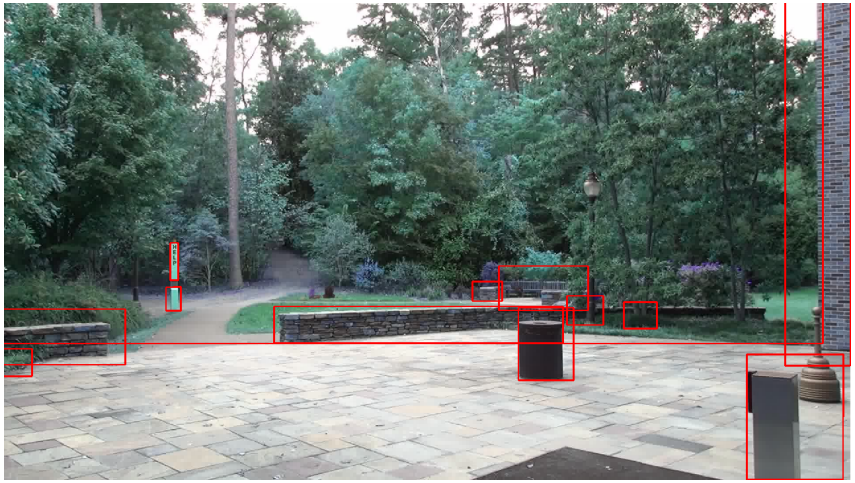
Object model construction I

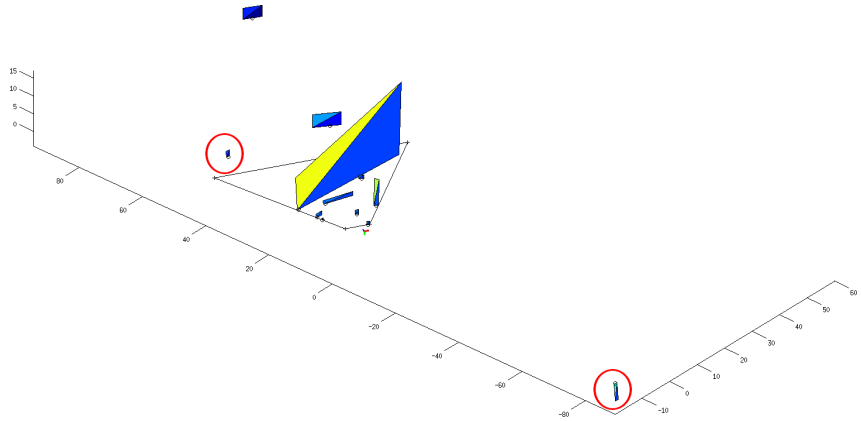




- For large, spread objects, the ground point assumption fails - see for example the bush areas on both image sides

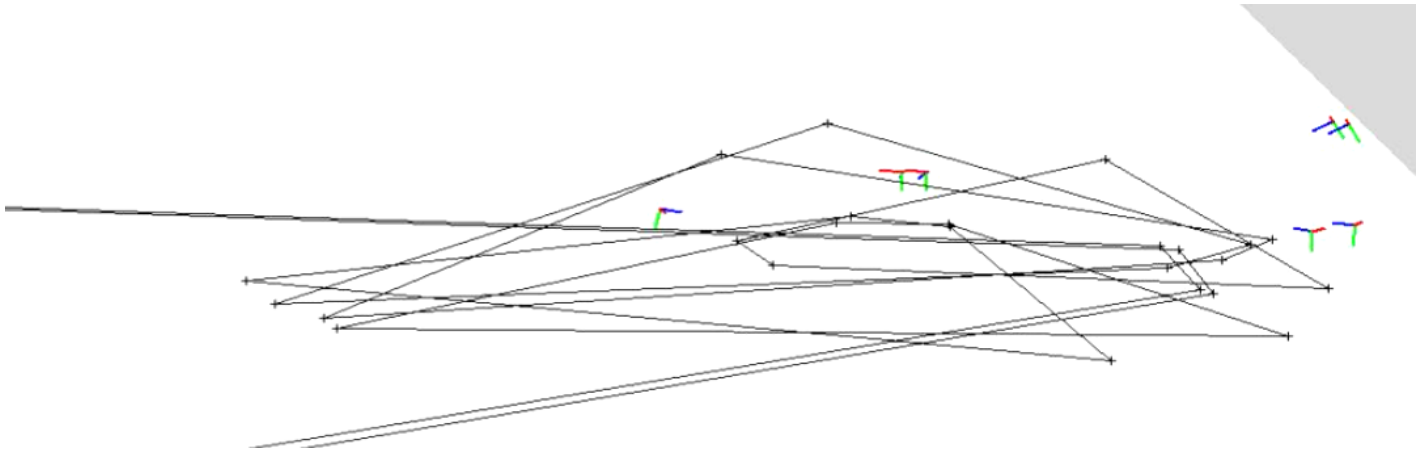
Object model construction I





- In this case, a partial segmentation of the object leads to a wrong reconstruction, because the begining of the object is “beyond the horizon”
- Multi-view reconstruction can help mitigate these errors

Scene coverage from cameras I



- Assuming a known horizon / preset horizon, and ground plane.
- We cropped the horizon at 100 meters – to avoid errors due to ground plane inaccuracies.
- Basically - allows us to reason about camera object coverage and utilize resources efficiently.

Uncertainty in multiview scene understanding - tracking

Standard methods for multi-view, multi-target tracking often reason about how to use 3D/multiview information

- Tracking-Reconstruction - first do multitarget 2D tracking, then associate to form 3D tracks.
- Reconstruction-Tracking - first do reconstruction of target candidates, that associate in 3D to form 3D trajectories.
- Combined approaches (3D+2D)- combine 3D reconstruction, image-based tracking, and data association.

3D scene model construction I

Conclusions / Future work:

- Closer integration with 3D reconstruction / multiview can help a lot - esp. for scene modelling and tracking.
- Match/tracking outliers, moving cameras, bad viewpoints, should be addressed, with uncertainty estimates.
- Ground planes provides good approximation for upright 3D objects, should be combined with multiview reasoning to avoid errors and provide better 3D modelling where available.
- 3D priors should be used to improve reasoning in 2D predicates, with proper uncertainty estimates.
- Incorporating scene analysis and classification into the 3D reconstruction can help.