Combination of Segmentation, Depth Estimation, and Priors

Helmut Mayer, Martin Drauschke & Jan Bartelsen
Institute of Applied Computer Science, Bundeswehr University Munich

Heiko Hirschmüller
Institute of Robotics and Mechatronics, DLR, Oberpfaffenhofen

Specification and Cause of the Problem

Given: One or more perspective images mostly from different view points of 3D objects such as buildings, trees, or faces

- Images can be **segmented** (comprises feature extraction from our point of view) or be used to **estimate depth** by stereo, shape-from-shading, shape-from-texture, or from silhouettes.
- Segmentation as well as depth estimation have **deficits** due to weak contrast, repetitive textures, shadows, occlusions, noise, etc.



Specification and Cause of the Problem

Types of image edges according to (BINFORD 1981)

- **G**eometrical limit of visible surface
- **I**llumination discontinuity
- **R**eflectivity boundary

Seven types of edges

- G, I, R (single reason)
- GI, GR, IR (two reasons)
- GIR (three reasons)

Specification and Cause of the Problem

Idea: Why not link segmentation and depth estimation, possibly including also the **semantics** of objects in the scene?

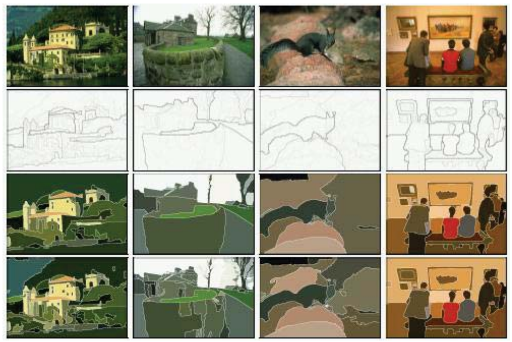
Two level chicken and egg problem

- 2D segmentation (edges, regions) could improve 3D (dense) depth estimation, while 3D information could improve segmentation.
- Segmentation (in 2D and 3D) could be used to infer object types, which in turn could be employed to improve segmentation and depth estimation by making available suitable priors for shape, size, or radiometric properties.

Table of Contents

Segmentation

- As demonstrated by (ARBELÁEZ et al. 2011) segmentation can benefit from multiple local cues, globalization, and a hierarchical region tree – image (top), Ultrametric Contour Map, optimal dataset scale, and optimal image scale (bottom) segmentations.



Depth Estimation

- Semi Global Matching (Hirschmüller 2008) – Berlin from 1,600 Microsoft UltraCam images with 86 Megapixels, 7 cm on ground

Depth Estimation

- High-end (according to Daniel Scharstein's test page) stereo matching (Hiep et al. 2009)

- Stereo and shape-from-shading, etc. (for fusion, e.g., (Fua & Leclerc 1995)) have reached a very high level.
- Yet, there are still **problems** with weakly and repetitively textured areas, at occlusion boundaries, with stronger perspective distortions, lighting changes, and complex reflection functions.

Segmentation and Depth Estimation

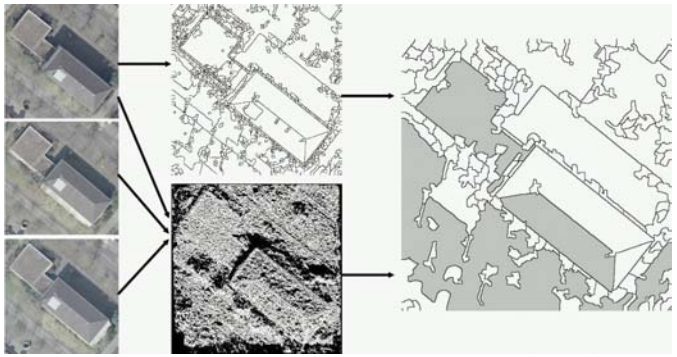
- **Idea:** Iterative procedure starting with independent segmentation and depth estimation.
- Segmentation can be used to avoid gaps in depth estimation, but also to delineate silhouettes.
- On the other hand, from depth estimation surfaces can be derived. Can help to improve the segmentation in the images.

Segmentation and Depth Estimation

- Segmenting palm leaves could avoid gaps in depth estimation.
- Approximating depth by surfaces, particularly planes, and extending them over small occlusions, depth estimation could be biased towards more meaningful results for weakly textured areas.
- **Which parts can one trust?** Is the segmentation and / or the 3D reconstruction reliable?
- **At which point is semantics needed?** I.e., long narrows strips versus certain types of (palm) leaves

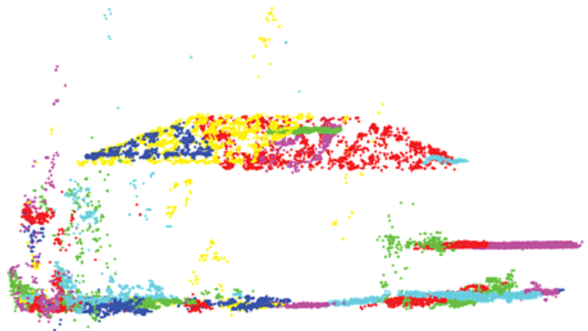
Segmentation and Depth Estimation

- Drauschke et al. (2009) segment reference image and estimate depth from multiple images.
- Adjacent segments are merged according to 3D information.



Segmentation and Depth Estimation

- Particularly, in (Drauschke et al. 2009) the best fitting plane is computed for all 3D points of each segment.
- Results reasonable for roof and ground
- Problems with vegetation (model does not fit)



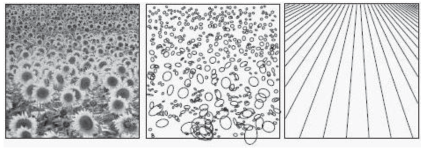
Segmentation and Depth Estimation

- SCHNABEL et al. (2007) show how not only planes (red in third column), but also cylinders (green), spheres (yellow), cones (purple), and tori (grey) can be derived via RANSAC.

- They could be used to support depth estimation beyond plane sweeping.

Segmentation and Depth Estimation

- Segmentation can be used to improve depth estimation based on **silhouettes**.
- On the other hand, known 3D structure can be used to infer and then eliminate **shading** effects leading to better segmentation results.
- While the deformation of texture can be used to infer depth (shape-from-texture (LINDEBERG 1994)), known depth can be used to eliminate the distortion of texture concerning projection angle and scaling leading to a better texture classification



Object Recognition and Priors

Inclusion of prior information in similar way as for the combination of segmentation and depth estimation:

- Characteristics of the texture or specific shapes are used to infer object types.
- Object types imply priors on shape, depth structure, texture, etc.
- The priors will in turn lead to an improved 2D and 3D structure either confirming or contradicting the characteristics leading to the prior.

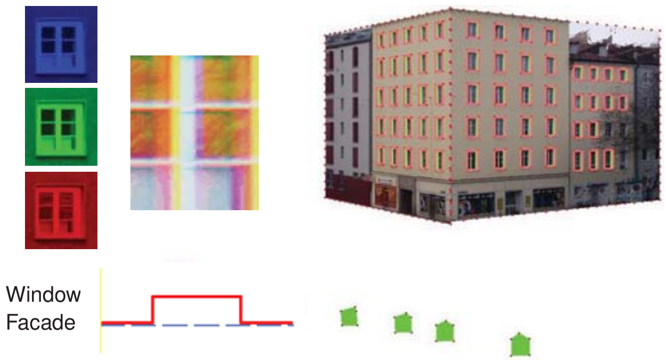
Object Recognition and Priors

Depth estimation for rows / columns of windows

- In (REZNIK & MAYER 2007) windows are delineated via an Implicit Shape Model (LEIBE & SCHIELE 2004).
- Organized in rows or columns balancing fit to data and simplicity of model via Akaike Information Criterion – AIC (AKAIKE 1973)

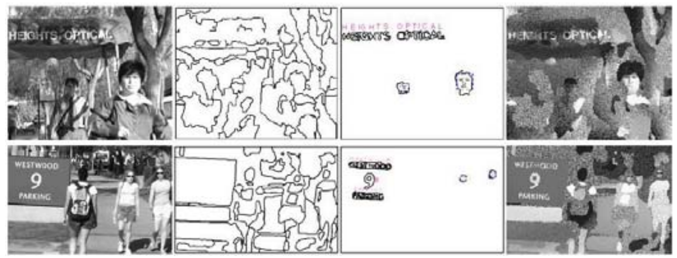
Object Recognition and Priors

- Knowledge about rows or columns of windows exploited when estimating their depth via **plane sweeping** for multiple images by computing optimum depth for whole row or column.



Object Recognition and Priors

- (TU & ZHU 2002) shows that "segmentation" of scene can be augmented by recognition for specific objects such as faces or characters



- Could be employed to bias depth reconstruction towards the typical 3D shape of a face.

Object Recognition and Priors

- How good needs to be the (2D / 3D) segmentation so that we can infer the semantics reliably?
- How much semantics needed for reliable segmentation or depth estimation? Are priors derived from, e.g., texture / color or certain object shapes also appropriate?

Accuracy and Reliability

- Combination of information should consider (geometrical and semantical) accuracy as well as reliability. Information might be estimated to be highly accurate, but is still unreliable.
- Ideally, conditions under which unreliability arise are known, can be detected, and there is another means which is reliable under these conditions. E.g., GPS is global solution, but unreliable if satellites occluded. INS (Inertial) only short time stable, but independent.
- **Idea:** Model / learn about accuracy / reliability of depth estimation, segmentation, etc.

Conclusion

- Combining segmentation and depth estimation has potential to improve both, particularly if semantics is considered.
- This could be the basis for a deeper modeling including the classification of materials (e.g., use of specific BRDFs) and the estimation of scene lighting (possibly including radiosity).
- For moving objects optical flow needs to be included.
- **How important is semantics**? How much semantics needed?
  - Prior can be obviously semantic such as “rows of windows”.
  - On the other hand, the texture of leafs can lead to a prior favoring small, approximately planar, elongated and rounded surfaces, without explicit knowledge about leaves or trees.
- **Model / learn accuracy and reliability!**
- Combined evaluation “Middlebury style”?

Literature I

AKAIKE, H. (1973): Information Theroy and an Extension of the Maximum Likelihood Principle, *Second International Symposium on Information Theory*, 267–281.

ARBELÁEZ, P., MAIRE, M., FOWLKES, C. & MALIK, J. (2011): Contour Detection and Hierarchical Image Segmentation, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **33**(5): 898–916.

BINFORD, T. (1981): Inferring Surfaces from Images, *Artificial Intelligence* **17**(1-3): 205–244.

DRAUSCHKE, M., ROSCHER, R., LÄBE, T. & FÖRSTNER, W. (2009): Improving Image Segmentation Using Multiple View Analysis, *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Volume (38) 3/W4, 211–216.

FUA, P. & LECLERC, Y. (1995): Object-Centered Surface Reconstruction: Combining Multi-Image Stereo and Shading, *International Journal of Computer Vision* **16**(1): 35–56.

HIEP, V., KERIVEN, R., LABATUT, P. & PONS, J.-P. (2009): Towards High-resolution Large-scale Multi-view Stereo, *Computer Vision and Pattern Recognition*, 1430–1437.

HIRSCHMÜLLER, H. (2008): Stereo Processing by Semi-Global Matching and Mutual Information, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **30**(2): 328–341.

Literature II

LEIBE, B. & SCHIELE, B. (2004): Scale-Invariant Object Categorization Using a Scale-Adaptive Mean-Shift Search, *Pattern Recognition – DAGM 2004*, Springer-Verlag, Berlin, Germany, 145–153.

LINDEBERG, T. (1994): *Scale-Space Theory in Computer Vision*, Kluwer Academic Publishers, Boston, USA.

REZNIK, S. & MAYER, H. (2007): Implicit Shape Models, Model Selection, and Plane Sweeping for 3D Facade Interpretation, *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Volume (36) 3/W49A, 173–178.

SCHNABEL, R., WAHL, R. & KLEIN, R. (2007): Efficient RANSAC for Point-Cloud Shape Detection, *Computer Graphics Forum* **26**(2): 214–226.

TU, Z. & ZHU, S. (2002): Image Segmentation by Data-Driven Markov Chain Monte Carlo, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **24**(5): 657–673.