3D Model Assisted Image Segmentation

Srimal Jayawardena   Di Yang   Marcus Hutter
Australian National University

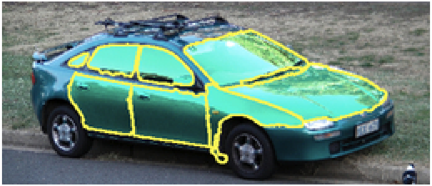
srimal(dot)jayawardena(at)anu(dot)edu(dot)au
http://users.cecs.anu.edu.au/~srimalj

DICTA 2011

The problem

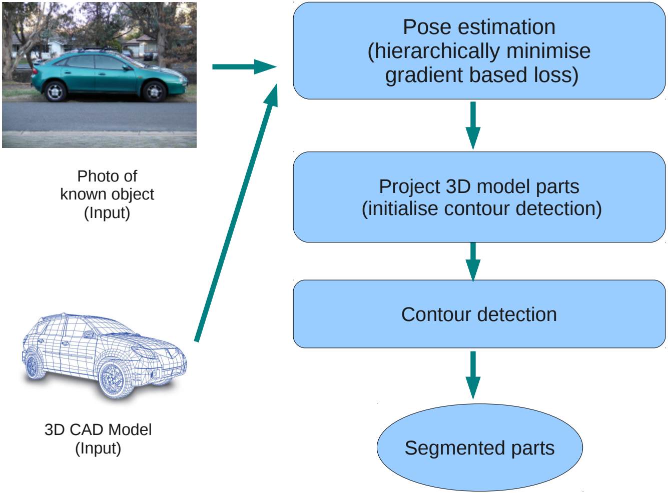
Segmenting a mostly homogeneous (same color/texture) object into parts is a hard problem.

(a) Original Image



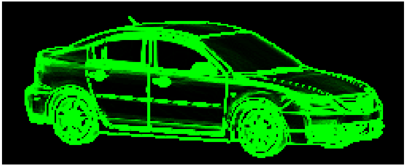
(b) Segmentated into parts

Methodology Overview



Gradient Loss for Pose Estimation

Let θ parameterize the pose of the 3D model w.r.t the camera.



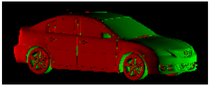
(a) 3D Model Gradients $\boldsymbol{G_N(\theta)}$

(b) Photo Gradients $\boldsymbol{G_I}$

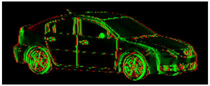
Loss at pose $\boldsymbol{\theta}$,

$$L_g(\boldsymbol{\theta}) := 1 - (corr(\boldsymbol{G_N(\theta)}, \boldsymbol{G_I}))^2 \in [0, 1]$$

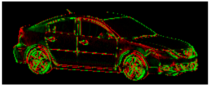
3D Model Gradients



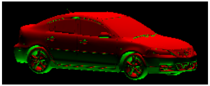
(a) $\Phi_x(u, v, \boldsymbol{\theta})$



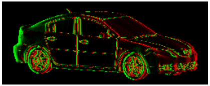
(b) $\frac{\partial \Phi_x(u,v,\boldsymbol{\theta})}{\partial u}$



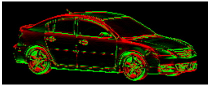
(c) $\frac{\partial \Phi_x(u,v,\boldsymbol{\theta})}{\partial v}$



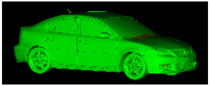
(d) $\Phi_y(u, v, \boldsymbol{\theta})$



(e) $\frac{\partial \Phi_y(u,v,\boldsymbol{\theta})}{\partial u}$

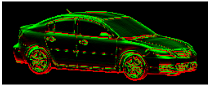


(f) $\frac{\partial \Phi_y(u,v,\boldsymbol{\theta})}{\partial v}$

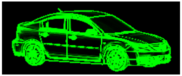


(g) $\Phi_z(u, v, \boldsymbol{\theta})$

(h) $\frac{\partial \Phi_z(u,v,\boldsymbol{\theta})}{\partial u}$



(i) $\frac{\partial \Phi_z(u,v,\boldsymbol{\theta})}{\partial v}$



(j) $\boldsymbol{G_N(\theta)}$

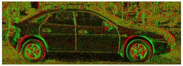
$$G_N(\boldsymbol{\theta})(u, v) = ||\nabla\Phi(u, v, \boldsymbol{\theta})||_k^k \qquad (1)$$

Photo Gradients

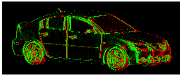
(a) Real photo

(b) Synthetic

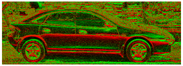
Using grayscale intensity



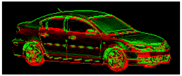
(c) Real $\frac{\partial \boldsymbol{I}}{\partial u}$



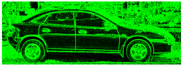
(d) Synthetic $\frac{\partial \boldsymbol{I}}{\partial u}$



(e) Real $\frac{\partial \boldsymbol{I}}{\partial v}$



(f) Synthetic $\frac{\partial \boldsymbol{I}}{\partial v}$)



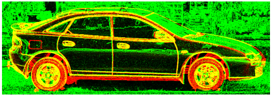
(g) Real $\boldsymbol{G_I}$

(h) Synthetic $\boldsymbol{G_I}$

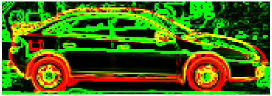
$$G_I(u,v) = ||\nabla I(u,v)||_k^k \qquad (2)$$

Overlays and Smoothing

(a) Real

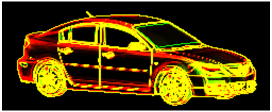


(b) n=0

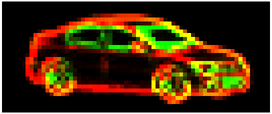


(c) n=2

(d) Synthetic

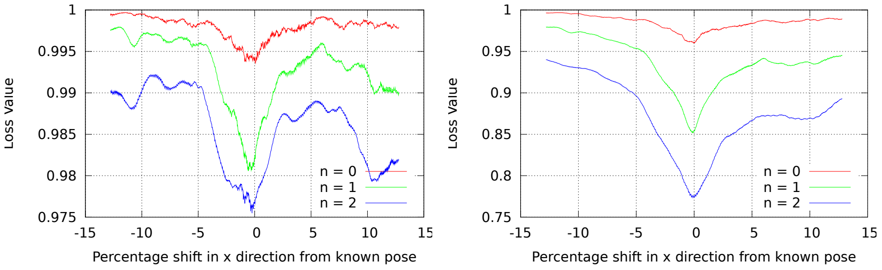


(e) n=0



(f) n=2

Loss Landscapes



(a) 2-norm (b) 1-norm

Hierarchical Optimization

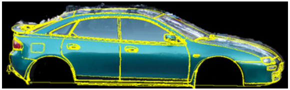
(a) Photo

(b) Background removed

(c) n=2

(d) n=1

(e) n=0



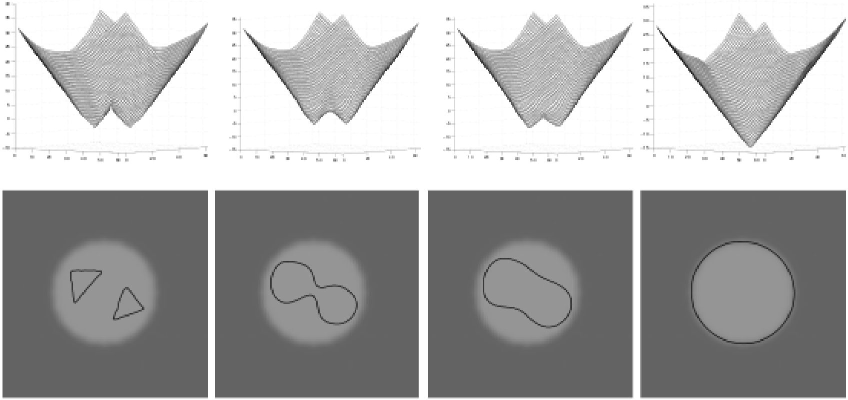
(f) Final fine pose n=0

Next: Initialise a *Level Set Evolution* contour detection from projected 3D model parts

Contour Detection

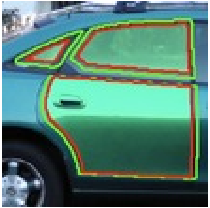
Level Set Evolution without re-initialization [Li et al., 2005, CVPR]

Row 1: Level set function, Row 2: Zero level curve

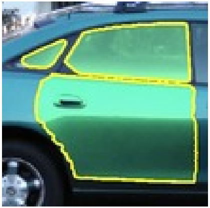


(a) Initialisation (b) (c) (d) Final

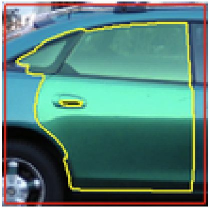
Results



(a) Initialisation



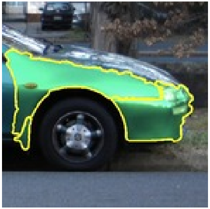
(b) Result



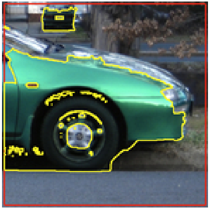
(c) Benchmark GC

(d) Benchmark LS

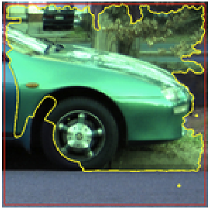
(e) Initialisation



(f) Result

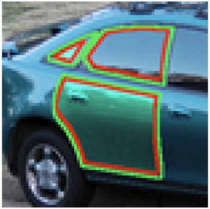


(g) Benchmark GC



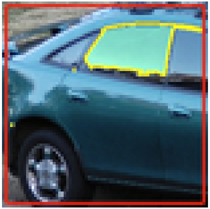
(h) Benchmark LS

Results

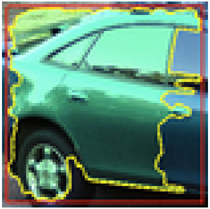


(a) Initialisation

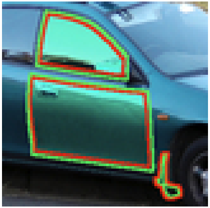
(b) Result



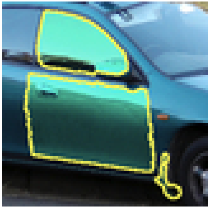
(c) Benchmark GC



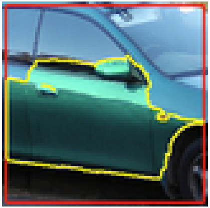
(d) Benchmark LS



(e) Initialisation



(f) Result



(g) Benchmark GC

(h) Benchmark LS

Accuracy

- Part segmentation results for two views of a Mazda Astina.
- $Accuracy = 1 - \left(\frac{No.Misclassified.Pixels}{No.Ground.Truth.Pixels} \right)$

Part	Side View	Semi Profile	Avg.
Fender	97.7%	97.6%	97.7%
Front door	98.1%	95.3%	96.7%
Back door	96.8%	93.6%	95.2%
Mud flap	97.3%	95.1%	96.2%
Front window	97.8%	97.5%	97.7%
Back window	99.5%	93.9%	96.7%

Discussion

- Challenges - High amount of reflections and noise
- A closer initialisation curve - better results
- Future work - simlutaneous pose estimation and segmentation

Thank you!

References I

Li, C., Xu, C., Gui, C., and Fox, M. (2005).

Level set evolution without re-initialization: a new variational formulation.

In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 1, pages 430 – 436 vol. 1.