Occlusion Robust Pose and Model Estimation using Gaussian Process Latent Variable Model on GPU

Christopher B. Choy

Electrical Engineering, Stanford University

Objectives

- Given **monocular image/video** and **collection of 3D models**
- Find :
  - **Pose Estimation** : 3D location, 3D rotation 1D scale
  - **Model variation estimation** : 3D CAD, Skeletal or Voxelized model
- s.t. 3D model fits foreground segmentation

Pipeline

- Foreground background segmentation
- Model variation using GPLVM
- Energy Function and Optimization
- Multiple Objects using non-intersection constraints

Segmentation

Object Detection using Hierarchical Segmentation

- Edges from Garbor filters
- Graph over the image using the edges
- n smallest eigenvectors of the graph Laplacian as new input images.

$$W_{ij} = \exp\left(-\max_{p \in \overline{ij}} \frac{mPb(p)}{\rho}\right) \text{ for } ||p_i - p_j|| < r \quad (1)$$

$$\mathbf{L} = \mathbf{D} - \mathbf{W} \text{ where } \mathbf{W} = [W_{ij}]_{ij} \text{ and } \mathbf{D}_{jj} = \sum_i W_{ij} \quad (2)$$

- n smallest eigenvalues of $\mathbf{L}$ as input for contour detection $\rightarrow sPb(p)$
- linearly combine $sPb(p)$ and $mPb(p)$ to get global contour.
- Oriented watershed : averages the edges on the same line segment
- Merges the regions with the smallest average weight of the edge in common

Gaussian Process Latent Variable Model

Probabilistic PCA $\rightarrow$ Dual PPCA $\rightarrow$ Non-linear Dual PPCA (GPLVM)

- $y = Wx + \eta$, Noise : $\eta = \mathcal{N}(\eta|\mathbf{0}, \beta^{-1}\mathbf{I})$

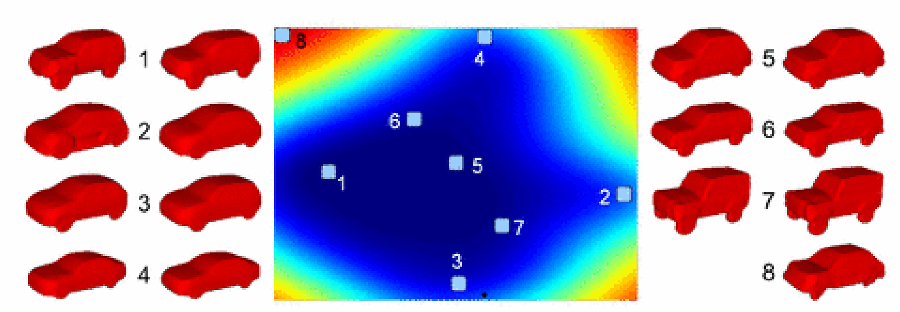
$$L = -\frac{DN}{2} ln(2\pi) - \frac{D}{2} ln|K| - \frac{1}{2} tr(K^{-1}Y^TY) \quad (3)$$

$$\text{where } K = WW^T + \beta^{-1}\mathbf{I}$$

$$L = -\frac{DN}{2} ln(2\pi) - \frac{D}{2} ln|K| - \frac{1}{2} tr(K^{-1}YY^T) \quad (4)$$

$$\text{where } K = XX^T + \beta^{-1}\mathbf{I}$$

- dual PPCA $X = ULV^T$ first q eigen.. of $\frac{1}{D}YY^T$
- PPCA $W = U'L'V'^T$ first q eigen.. of $\frac{1}{N}Y^TY$
- Arbitrary nonlinear Kernel: $\kappa(x_i, x_j) = \beta_1 exp\left(\frac{||x_i - x_j||^2}{2\beta_2}\right) + \beta_3 + \beta_4 \delta_{ij}$
- DCT Compression $20 \times 20 \times 20 = 8k$
- $Y \in \mathcal{R}^{n \times D}$ where n is the number of model



Borrowed from A. Dame, et al. CVPR 2013

Energy Function and Newton Step

$$E(\Phi) = \sum_{x \in \Omega} log[\pi(\Phi(x))P_f(x) + (1 - \pi(\Phi(x)))P_b(x)]$$

Where P_f foreground posterior probability

$$\frac{\partial E}{\partial \lambda_p} = \sum_{x \in \Omega} \frac{P_f(x) - P_b(x)}{\pi(\Phi)P_f(x) + (1 - \pi(\Phi))P_b(x)} \frac{\partial \pi(\Phi)}{\partial \lambda_p}$$

$$\frac{\partial \pi(\Phi)}{\partial \lambda_p} = (1 - \pi(\Phi)) \sum_z \frac{e^{\Phi(X)}}{e^{\Phi(X)} + 1} \frac{\partial \Phi}{\partial X} \frac{\partial X}{\partial \lambda_p}$$

- Analytic solution $\frac{\partial X}{\partial \lambda_p}$ exists
- $\frac{\partial \Phi}{\partial X}$ using centered finite differences
- x pixel on image, X_{camera} voxelize 3D space.
- $X_{camera} = {}^vM_o X_{object}$ where ${}^vM_o$ is transformation matrix

Multi-object Pose Estimation

- Multi object : occlusion
- Occlusion robust pose model estimation
- Constraints: non-intersection

$$\min_\lambda \sum_{o \in \mathbf{O}} \sum_{x \in \Omega} log[\pi(\Phi^o(x))P_f^o(x) + (1 - \pi(\Phi^o(x)))P_b^o(x)]$$

$$\text{s.t. } \sum_{o \in \mathbf{O}} \pi(\Phi_o) \leq 1$$

- Optimization too high dimensional i.e. $\lambda \in \mathcal{R}^{|\mathbf{O}|(7 + dim\mathcal{L})}$
- Alternative optimization

while convergence

for $o \in \mathbf{O}$

$$\min_\lambda \sum_{x \in \Omega} log[\pi(\Phi^o(x))P_f^o(x) + (1 - \pi(\Phi^o(x)))P_b^o(x)]$$

$$\text{s.t. } \pi(\Phi_o) \leq \sum_{o\prime \in \mathbf{O} \setminus \{o\}} \pi(\Phi_{o\prime})$$

- $\lambda \in \mathcal{R}^{(7 + dim\mathcal{L})}$, $|\mathbf{O}|$ times
- Non-intersection$\rightarrow$ background for other objects

while convergence

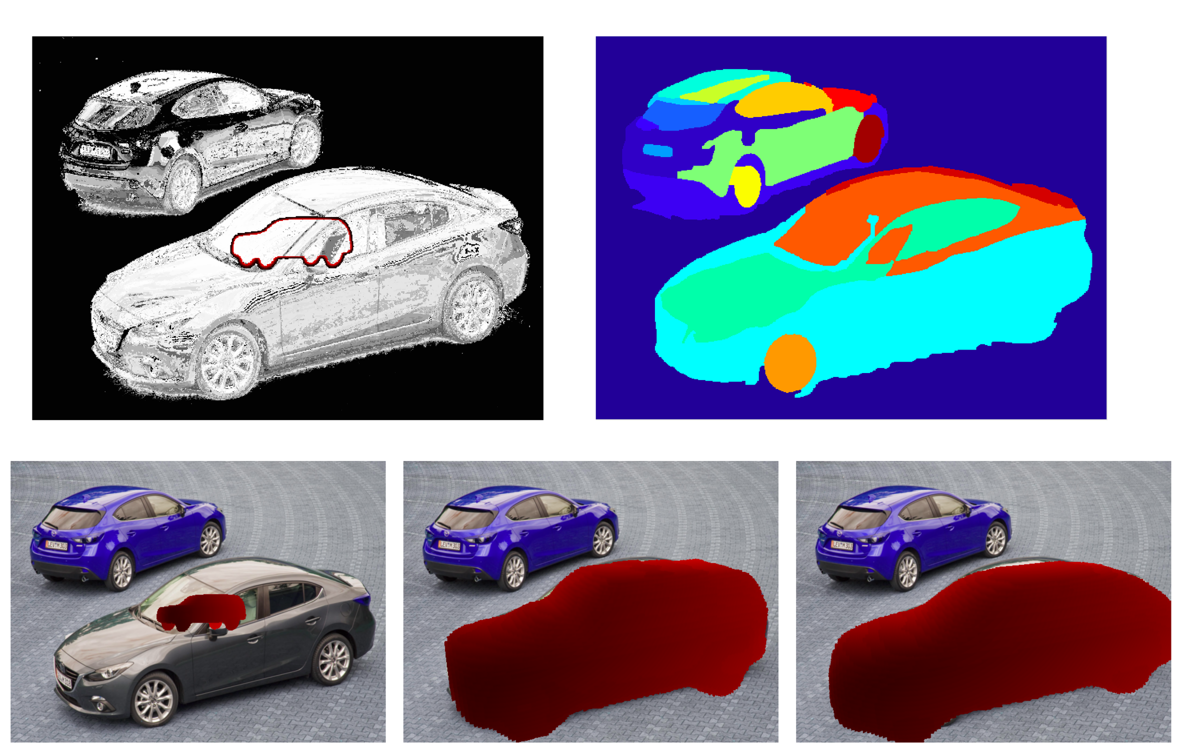
for $o \in \mathbf{O}$

$$P_b^o(x) \leftarrow \mu\pi(\Phi^{o\prime}) \quad P_f^o(x) \leftarrow 1 - \mu\pi(\Phi^{o\prime})$$

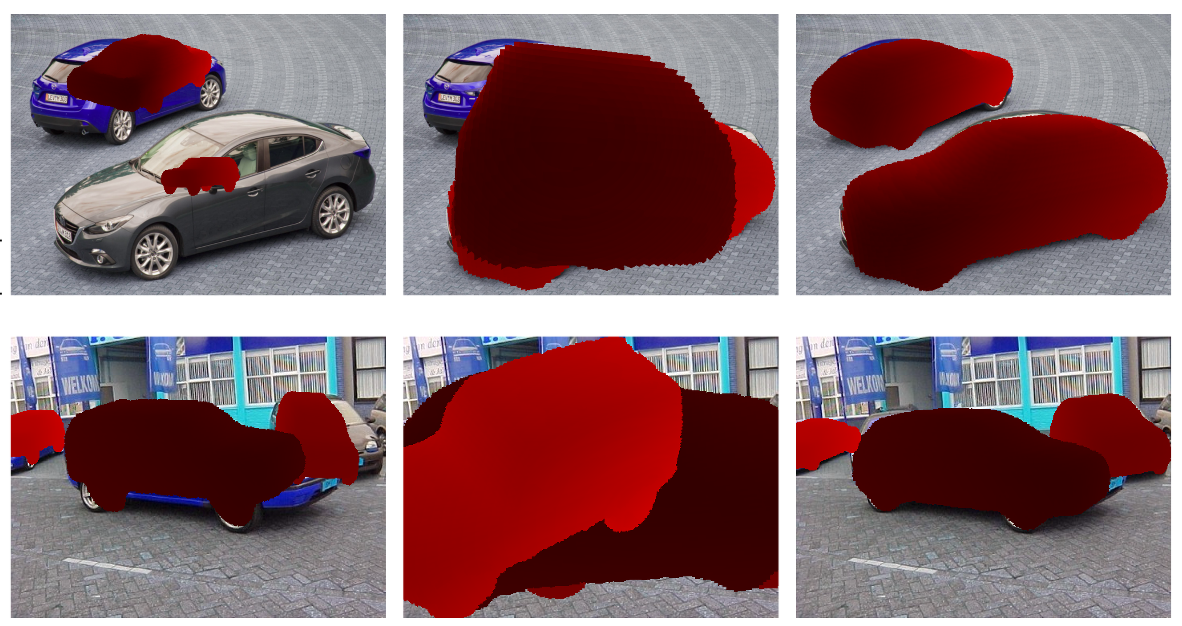
$$\min_\lambda \sum_{x \in \Omega} log[\pi(\Phi^o(x))P_f^o(x) + (1 - \pi(\Phi^o(x)))P_b^o(x)]$$

- Using only the non-occluded parts to estimate the pose $.7 \geq \mu \geq .5$

Results



Results



Conclusion

- Occluding object affect the pose of the occluded object
- Estimate better pose through object relationship

Discussion

- Evaluation metric
  - segmentation overlap : VOC 2012 segmentation
  - pose ground truth label and L2 distance as error metric
- Automatic initialization
  - DPM-style rough pose estimation
  - Better segmentation for complicated scene
- 3-D non-intersection constraint using B.B.
- Dynamic segmentation using 3D model i.e. dynamic graph cut
- Supporting plane $\rightarrow$ depth and scale

References

[1] A. Dame, V. A. Prisacariu, C. Y. Ren and I. D. Reid, "Dense Reconstruction Using 3D Object Shape Priors". CVPR, 2013

[2] Felzenszwalb, P. F. and Girshick, R. B. and McAllester, D. and Ramanan, D. "Object Detection with Discriminatively Trained Part Based Models", PAMI 2010