Occlusion Robust Pose and Model Estimation using GPLVM on GPU

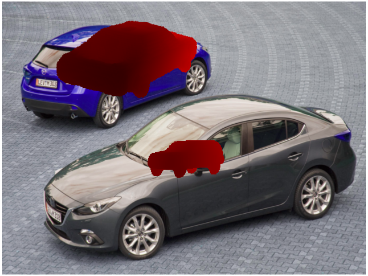
Christopher B. Choy

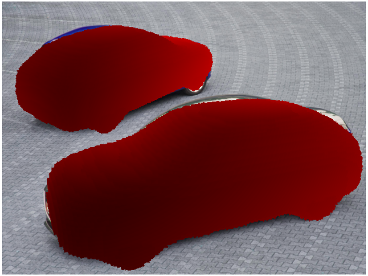
Stanford University

chrischoy@stanford.edu

January 17, 2014

The Table of Contents





Pose and Model Variation

- Given a monocular image/video and a collection of 3D models
- Find
  - Pose Estimation : 3D location, 3D rotation 1D scale
  - Rigid transformation
  - Model variation estimation : 3D CAD, Skeletal or Voxelized model
- such that the 3D model fits data (foreground mask) well

Recent Progress

- 2D static model
  - Riklin-Raviv, T., et al. (2007). Prior-based Segmentation and Shape Registration in the Presence of Perspective Distortion. IJCV
- 3D static model
  - Schmaltz, C., et al. (2007). Region-Based Pose Tracking. PAMI
- 3D dynamic model
  - Kohli, P., et al. (2008). Simultaneous Segmentation and Pose Estimation of Humans Using Dynamic Graph Cuts. IJCV
  - Dambreville, S., et al. (2008). Robust 3d pose estimation and efficient 2d region-based segmentation from a 3d shape prior. ECCV
  - Sandhu, R., et al. (2009). Non-Rigid 2D-3D Pose Estimation and 2D Image Segmentation. CVPR
  - Prisacariu, V., et al. (2012). PWP3D: Real-Time Segmentation and Tracking of 3D Objects. IJCV
  - Dame, A., et al. (2013). Dense Reconstruction Using 3D Object Shape Priors, CVPR

Motivation and Main Contribution

- Recent works on the pose estimation of an object
  - not occlusion robust
  - multiple objects?

- Work based on Dame et al. CVPR 2013
- Automatic segmentation
- Better pixel statistics
- Multiple objects
- Using other objects to find pose that are robust to occlusion

Energy Function

- Measure the goodness-of-fit (pose and model variation)
- Overlap between the silhouette and the projection foreground, the complement of sihouette and the projection background.

$$E(\Phi) = \int_\Omega [\pi(\Phi)P_f + (1 - \pi(\Phi))P_b] \, d\Omega$$

- Discretize, x pixel on the view plane, X_{camera} voxelize 3D space.
- $X_{camera} = {}^vM_o X_{object}$ where ${}^vM_o$ is a transformation matrix
- $P_f(x)$ posterior probability of an image, $\Phi(X_c)$ the SDF of the voxel X_c.
- SDF: negative distance from the model, (positive inside)
- $\pi(\Phi)$ is the projection function $1 - \prod(1 - \frac{1}{1+e^{-\Phi(X)}})$ i.e. noisy OR

Newton Step

- minimize the objective function for 7D rotation + translation + scale
- Search over all models $\rightarrow$ Discrete space! Inefficient
- Continuous relaxation?
- Gradient descent for the model variation
    - Low dimensional representation

Variations of PCA

- Dimensionality Reduction technique: PCA!
- Which variation?
- Probabilistic PCA $\rightarrow$ Dual PPCA $\rightarrow$ Non-linear Dual PPCA (GPLVM)

Probabilistic PCA

- $y = Wx + \eta$, Data: centered $y \in \mathcal{R}^D$, Latent variable : $x \in \mathcal{R}^q$, Mapping: $W \in \mathcal{R}^{D \times q}$ and Noise : $\eta = \mathcal{N}(\eta | \mathbf{0}, \beta^{-1}\mathbf{I})$
- $p(y|x, W, \beta) = \mathcal{N}(y|Wx, \beta^{-1}\mathbf{I})$
- The red pill or the blue pill: Marginalize either W or x
  The blue pill : $p(x) = \mathcal{N}(x|\mathbf{0}, \mathbf{I})$
- Then we marginalize over the latent variable
  $p(y|W, \beta) = \mathcal{N}(\mathbf{0}, WW^T + \beta^{-1}\mathbf{I})$
- Log likelihood L

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

where $K = WW^T + \beta^{-1}\mathbf{I}$ seem familiar? $\rightarrow$ Kernel

- $\nabla_W L = 0$ then $W = U' L' V'^T$ where U' is the first q eigenvectors of $\frac{1}{N} Y^T Y$ (Covar) $L_{ii} = (\Lambda_i + 1/\beta)^{-1/2}$

Dual Probabilistic PCA

- The red pill : $P(W) = \prod^D \mathcal{N}(w_i|\mathbf{0}, \mathbf{I})$

$$L = -\frac{DN}{2} ln(2\pi) - \frac{D}{2} ln|K| - \frac{1}{2} tr(K^{-1}YY^T) \tag{3}$$

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

- $\nabla_X L = 0$
- dual PPCA $X = ULV^T$ where U first q eigenvectors of of $\frac{1}{D}YY^T$
- PPCA $W = U'L'V'^T$ where U' first q eigenvectors of $\frac{1}{N}Y^TY$
- Linear Kernel $\kappa(x_i, x_j) = x_i^T x_j + \frac{1}{\beta}\delta_{ij}$
- Arbitrary nonlinear Kernel: $\kappa(x_i, x_j) = \beta_1 exp\left(\frac{||x_i - x_j||^2}{2\beta_2}\right) + \beta_3 + \beta_4\delta_{ij}$
- Non-linear Dual PPCA $\rightarrow$ Gaussian Process Latent Variable Model

Dual Probabilistic PCA

- Probabilistic sampling
- ML sample: $\mu^* = \mu + Y^T \mathbf{K}^{-1} \kappa(x^*, X)$
  $\sigma^{*2} = \kappa(x^*, x^*) + \kappa(x^*, X)^T K^1 \kappa(x^*, X)$
- Stochastic Gradient Descent (A. Yao et al., NIPS2011)

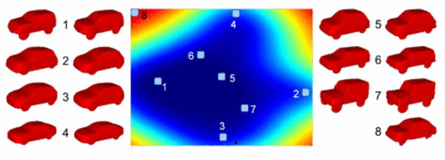
Algorithm 1: Stochastic GPLVM

Randomly initialize $\mathbf{X}$
Set $\boldsymbol{\beta}$ with an initial guess
for t = 1:T
  randomly select $\mathbf{x}_r$
  find R neighbors around $\mathbf{x}_r$: $\mathbf{X}_R = \mathbf{X} \in \mathcal{R}$
  Compute $\frac{\partial L}{\partial \mathbf{X}_R}$ and $\frac{\partial L}{\partial \boldsymbol{\beta}_R}$ (see Eq. (3))
  Update $\mathbf{X}$ and $\boldsymbol{\beta}$:
    $\Delta \mathbf{X}_t = \mu_X \cdot \Delta \mathbf{X}_{t-1} + \eta_X \cdot \frac{\partial L}{\partial \mathbf{X}_R}$
    $\mathbf{X}_t \leftarrow \mathbf{X}_{t-1} + \Delta \mathbf{X}_t$
    $\Delta \boldsymbol{\beta}_t = \mu_\beta \cdot \Delta \boldsymbol{\beta}_{t-1} + \eta_\beta \cdot \frac{\partial L}{\partial \boldsymbol{\beta}_R}$
    $\boldsymbol{\beta}_t \leftarrow \boldsymbol{\beta}_{t-1} + \Delta \boldsymbol{\beta}_t$
end

- Non-convex, Initialize using PCA
- Stochastic GPLVM Angela Yao et al.,
  `http://www.youtube.com/watch?v=1f_357NwdzQ`

Voxels to GPLVM

- Y is voxelized models
- Make a 3D grid of $125 \times 125 \times 125 \approx 2$million
- DCT of the grid, first 20 low freq components $D = 20 \times 20 \times 20 = 8k$
- Learn the latent space $\mathcal{X}$ from voxelized models
- $Y \in \mathcal{R}^{n \times D}$ where n is the number of model
- 2D cont. approximation of the space[1]



[1]Borrowed from A. Dame, et al. CVPR 2013

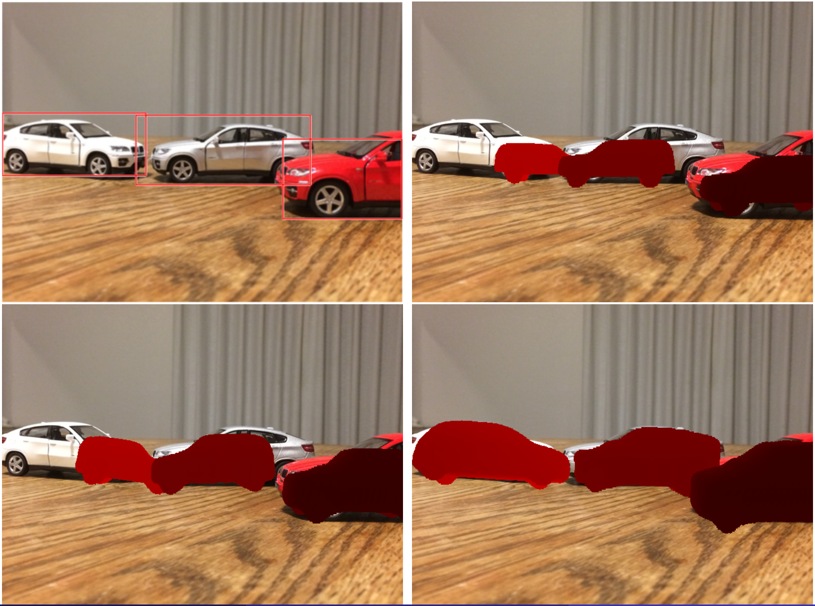
Foreground and Background segmentation

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

- Foreground posterior probability $\leftarrow$ color histogram
- Strongly affected by background color and texture
- Given foreground background color histograms, find the pixelwise posterior foreground probability[2]

[2]A. Dame et al, CVPR 2013

Key Contribution: DPM and GrabCut

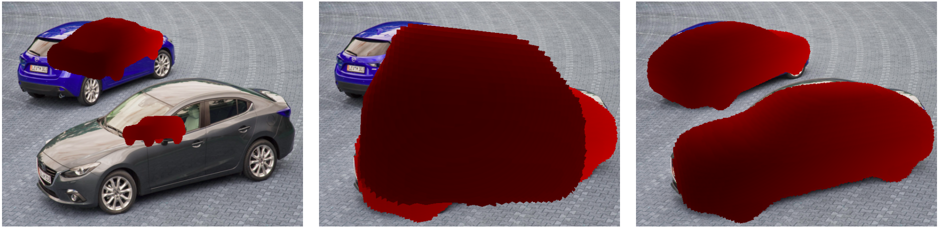


Gradient Step, step size 5×10^{-2} 5 frames per sec

Key Contribution: Multi object pose estimation

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

$$\begin{aligned} \underset{\lambda}{\text{minimize}} \quad & \sum_{o \in \mathbf{O}} \sum_{x \in \Omega} log[\pi(\Phi^o(x)) P_f^o(x) + (1 - \pi(\Phi^o(x))) P_b^o(x)] \\ \text{subject to} \quad & \sum_{o \in \mathbf{O}} \pi(\Phi_o) \leq 1 \end{aligned}$$



Key Contribution: Multi object pose estimation

- Optimization too high dimensional i.e. $\lambda \in \mathcal{R}^{|\mathbf{O}|(7+dim\mathcal{L})}$
- Alternative optimization: assuming that it's locally convex

  for $o \in \mathbf{O}$

$$\begin{aligned} \underset{\lambda}{\text{minimize}} \quad & \sum_{x \in \Omega} log[\pi(\Phi^o(x))P_f^o(x) + (1 - \pi(\Phi^o(x)))P_b^o(x)] \\ \text{subject to} \quad & \pi(\Phi_o) \leq \sum_{o\prime \in \mathbf{O} \setminus \{o\}} \pi(\Phi_{o\prime}) \end{aligned}$$

  end

  while convergence
- $\lambda \in \mathcal{R}^{(7+dim\mathcal{L})}$, $|\mathbf{O}|$ times
- Constraints too expensive i.e. $|\mathbf{O}| \times h \times w$
- Occlusion?

Key Contribution: Multi object pose estimation

- Constraint for non-intersection$\rightarrow$ background for other objects
- Posterior probabilities

  for $o \in \mathbf{O}$

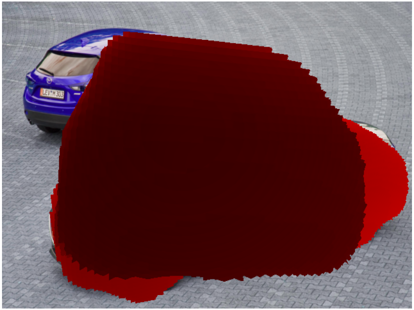
  $$P_b^o(x) \leftarrow \mu\pi(\Phi^{o'}) \qquad P_f^o(x) \leftarrow 1 - \mu\pi(\Phi^{o'})$$

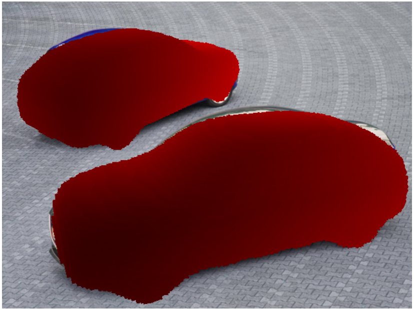
  $$\underset{\lambda}{\text{minimize}} \quad \sum_{x\in\Omega} log[\pi(\Phi^o(x))P_f^o(x) + (1 - \pi(\Phi^o(x)))P_b^o(x)]$$

  end

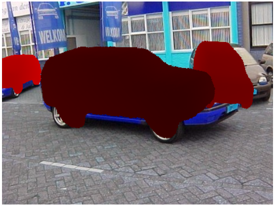
  while convergence
- Using only the non-occluded parts to estimate the pose $.7 \geq \mu \geq .5$

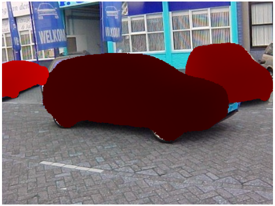
Gradient Step, step size 5×10^{-2} 5 frames per sec





Gradient Step, step size 5×10^{-2} 5 frames per sec





- Occluding object affect the pose of the occluded object
- Better pose estimation through object relationship

Conclusion

- Better pose estimation
  - Better segmentation
  - Object relationship
- Better evaluation metric and initialization

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
- Part based segmentation and part based 3D model
- **Virtual world construction from a monocular image**