3D Shape Representation

Tianqiang

04/01/2014
Why do we need 3D shapes?

- Image/video understanding
- Content creation
Why do we need 3D shapes?

- Image/video understanding

[Chen et al. 2010]
Why do we need 3D shapes?

- Image/video understanding

[Xiang et al. 2014 (ECCV)]

[Chen et al. 2010]
Why do we need 3D shapes?

Image/video understanding

Need a 3D shape prior!
Why do we need 3D shapes?

• Image/video understanding
• Content creation

[Yumer et al. 2010] [Kalogerakis et al. 2012]
Why do we need 3D shapes?

Content creation

Need a 3D shape prior, again!
Papers to be covered in this lecture

Recovering 3D from single images
• 3D Reconstruction from a Single Image [Rother et al. 08]
• Shared Shape Spaces [Prisacariu et al. 11]
• A Hypothesize and Bound Algorithm for Simultaneous Object Classification, Pose Estimation and 3D Reconstruction from a Single 2D Image [Rother et al. 11]
• Beyond PASCAL: A Benchmark for 3D Object Detection in the Wild [Xiang et al. 14 (WACV)]
• Inferring 3D Shapes and Deformations from Single Views [Chen et al. 10]
Papers to be covered in this lecture

Tracking

• Simultaneous Monocular 2D Segmentation, 3D Pose Recovery and 3D Reconstruction [Prisacariu et al. 12]
• Monocular Multiview Object Tracking with 3D Aspect Parts [Xiang et al. 14 (ECCV)]
Papers to be covered in this lecture

Shape synthesis

• A Probabilistic Model for Component-based Shape Synthesis [Kalogerakis et al. 12]
• Co-constrained Handles for Deformation in Shape Collections [Yumer et al. 14]
Recovering 3D from single images

Input: a single image
Recovering 3D from single images

Output: 3D shape, and pose or viewpoint
Outline

• A traditional generative model
• Shared shape space
• A benchmark for 3D object detection
Outline

• A traditional generative model
• Shared shape space
• A benchmark for 3D object detection
A Traditional Model

\[ P(S, V \mid I) \propto P(S, V, I) = P(I \mid S, V)P(S)P(V) \]

\( S \) : 3D shape (output)

\( V \) : Viewpoint (output)

\( I \) : Image (input)
A Traditional Model

\[ P(S, V | I) \propto P(S, V, I) = P(I | S, V)P(S)P(V) \]

- **S**: 3D shape (output)
- **V**: Viewpoint (output)
- **I**: Image (input)
A Traditional Model

\[ P(S, V \mid I) \propto P(S, V, I) = P(I \mid S, V)P(S)P(V) \]

Projection \hspace{1cm} Shape \hspace{1cm} Viewpoint

\text{prior} \hspace{1cm} \text{likelihood} \hspace{1cm} \text{prior}

[Rother et al. 2008]
[Xiang et al. 2014]
[Chen et al. 2014]
Rother et al. 2008 - Overview

- Voxel representation for 3D shape
- Viewpoint is known or enumerated
- Using an optics law to do projection
Rother et al. 2008 - Overview

1- 3D prior
2- 3D voxel occupancies
3- Projection
4- 2D pixel occupancies
5- Color model
6- Pixel color
Rother et al. 2008 – 3D Shape Prior

Voxel occupancies

(2D is just for illustration)
Rother et al. 2008 – Projection likelihood

(Assuming the viewpoint is known)

\[ E(0) = 1 - \varepsilon_0 \]
\[ E(1) = \varepsilon_1 \]

\[
P(Q_j = 0|V_1, \ldots, V_n) = \prod_{i=1}^{n} E(V_i),
\]
Two issues:
• Effect is unrelated to travel distance in each voxel
• Dependent on the resolution of grids
Rother et al. 2008 – Projection likelihood

Beer-lambert law:

\[ \frac{dl}{dr} = -l \cdot \alpha. \]

\[ \ln \frac{l}{l_0} = -\alpha \cdot r \Rightarrow P = \frac{l}{l_0} = e^{-\alpha \cdot r} \]

\[ P(Q_j = 0|V_1, \ldots, V_n) = \prod_{i=1}^{n} e^{-\alpha(V_i) \cdot r_{ji}} = \prod_{i=1}^{n} E(V_i)^{r_{ji}}, \]
Rother et al. 2008 – Optimization

Possible solution: belief propagation

But what if the graph is loopy?
Possible solution: belief propagation

But what if the graph is loopy?

Down-sample the image!
Rother et al. 2008 – Experiments

Learning shape prior: separately for each subclass
eg. Right leg in front, left leg up, ...

Degree = 0
Together

60

120
Separate

180
### Rother et al. 2008 – Experiments

<table>
<thead>
<tr>
<th>Unknown registration parameters</th>
<th>$E_R$ (%)</th>
<th>$V_R$ (%)</th>
<th>$E_\chi$ (cm)</th>
<th>$E_\phi$ (%)</th>
<th>$E_\theta$ (°)</th>
<th>$E_\rho$ (cm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>-</td>
<td>37.1</td>
<td>-1.8</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$\chi$</td>
<td>37.2</td>
<td>-2.2</td>
<td>6.2</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$\phi$</td>
<td>37.8</td>
<td>-2.2</td>
<td>-</td>
<td>57</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$\theta$</td>
<td>37.5</td>
<td>-2.3</td>
<td>-</td>
<td>-</td>
<td>45</td>
<td>-</td>
</tr>
<tr>
<td>$\rho$</td>
<td>36.4</td>
<td>-7.4</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>5.0</td>
</tr>
<tr>
<td>$\chi, \phi$</td>
<td>37.7</td>
<td>-2.9</td>
<td>8.5</td>
<td>48</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$\chi, \phi, \theta$</td>
<td>38.1</td>
<td>-3.6</td>
<td>12.4</td>
<td>81</td>
<td>65</td>
<td>-</td>
</tr>
</tbody>
</table>
Chen et al. 2010 - Overview

- Modeling intrinsic shape (phenotype) and pose variations
- Mesh representation for 3D shape.
- Only leveraging silhouettes in images
Chen et al. 2010 - Overview

Latent variables

$\mathcal{M}_S$

$x_S$

$\mathcal{M}_A$

$x_A$

Pose

Shape

Viewpoint

$V$

$\gamma_k$

$W_k$

$S_k$

Projection

$k = 1, 2, \ldots, K$

Intrinsic shape
Chen et al. 2010 - Overview

Latent variables

Pose

Viewpoint

Intrinsic shape

Shape

Projection

Silhouette

Latent variables:

$A_X$

$M_A$

$x_S$

$M_S$

$u$

$W_k$

$S_k$

$k = 1, 2, \ldots$
Chen et al. 2010 – 3D Shape Prior

Gaussian Process Latent Variable Models

Shape Generator

Pose Generator: Running
Gaussian Process Latent Variable Model

\[ P(u|x_A, \mathcal{M}_A) \]

\[ = \mathcal{N}(u; k_U^T (x_A) K_U^{-1} Y_A, (k_U(x_A, x_A) - k_U^T (x_A) K_U^{-1} k_U(x_A)) I) \]

\[ = \mathcal{N}(u; \bar{u}(x_A), \sigma_A^2(x_A) I) \quad (2) \]
Gaussian Process Latent Variable Model

Intrinsic shape (phenotype)

\[
P(v | x_S, M_S) = \mathcal{N}(v; k_V(x_S)K_V^{-1}Y_S, (k_V(x_S, x_S) - k_V^T(x_S)K_V^{-1}k_V(x_S))I)
= \mathcal{N}(v; \bar{v}(x_S), \sigma^2_S(x_S)I).
\]
\[ \Delta V'_i = J_i \Delta V_i. \]

\[ J_i = [m_i^{A,1}, m_i^{A,2}, m_i^{A,3}] [m_i^{O,1}, m_i^{O,2}, m_i^{O,3}]^{-1} \]
Chen et al. 2010 – Projection

\[ P(S_k \mid V, \gamma_k) = P(W_k \mid V, \gamma_k)P(S_k \mid W_k) \]

- \( S_k \): input silhouettes
- \( W_k \): projected silhouettes
- \( V \): samples on 3D mesh
- \( \gamma_k \): camera parameters
Chen et al. 2010 – Projection

\[
P(S_k \mid V, \gamma_k) = P(W_k \mid V, \gamma_k)P(S_k \mid W_k)
\]
Chen et al. 2010 – Projection

Projecting:

\[ P(W_k | V, \gamma_k) = \mathcal{N}(W_k; \tilde{P}_k V + \tilde{t}_k, \sigma^2_w I), \]

\[ \tilde{P}_k \quad : \text{projection} \]

\[ \tilde{t}_k \quad : \text{translation vector} \]
Chen et al. 2010 – Projection

Matching:

\[ P(S_k|W_k) = \frac{1}{Z} \exp \left( - \frac{1}{2\sigma_s^2} DT_{S_k}^2(W_k) \right) \]

\( DT \) : a similarity function

\( Z \) : normalizer
Chen et al. 2010 – Optimization

Approximation

\[
P(\{S_k\}_{k=1}^K | x_A, x_S, M_A, M_S, \{\gamma_k\}_{k=1}^K) \geq Q(x_A, x_S, \{\gamma_k\}_{k=1}^K)
\]

\[
= \prod_{k=1}^{K} \frac{1}{Z_k \sqrt{\det(I + \frac{1}{\sigma_s^2} \Sigma_{W_k})}} \exp( - \frac{1}{2\sigma_s^2} DT_{S_k}^2(\mu_{W_k}) )
\]

\[
(x_{A,ML}, x_{S,ML}, \{\gamma_{k,ML}\}_{k=1}^K) \approx \arg\min_{x_A, x_S, \{\gamma_k\}_{k=1}^K} - \log Q(x_A, x_S, \{\gamma_k\}_{k=1}^K).
\]

- Adaptive-scale line search
- Multiple initialization
Chen et al. 2010 – Experiments

<table>
<thead>
<tr>
<th>Data</th>
<th>Precision</th>
<th>Recall</th>
</tr>
</thead>
<tbody>
<tr>
<td>22 sharks</td>
<td>0.8996 ± 0.0481</td>
<td>0.9308 ± 0.0380</td>
</tr>
<tr>
<td>20 human bodies</td>
<td>0.7801 ± 0.0689</td>
<td>0.8952 ± 0.0995</td>
</tr>
</tbody>
</table>
Outline

• A traditional generative model
• Shared shape space
• A benchmark for 3D object detection
Shared Shape Space

Silhouette 3D shape

Latent variable

$X_{YZ}$

$Y$  $Z$
Shared Shape Space

Multimodality
Shared Shape Space

Silhouette 3D shape

\[ X_{YZ} \]

Latent variable

Cannot model multimodality
Shared Shape Space

3D shape  Silhouette  Similarity distribution

Latent variables

Can model multimodality
Shared Shape Space

$X_{YS}$  $X_{YZ}$  $y$  $S$
Shared Shape Space

A little more detail:
• Hierarchy of latent manifolds
• Representation of 2D and 3D shapes
Hierarchy of Manifolds

10D

\( X^0_{YZ} \)

4D

\( X^{r-1}_{YZ} \)

2D

\( X^r_{YZ} \)

Y

Z
2D/3D Discrete Cosine Transforms

Advantages:
• Fewer dimensions than directly representing the images.
2D/3D Discrete Cosine Transforms

Image:
640X480 = 307200 dims

DCT:
35X35 = 1225 dims
Shared Shape Space - Inference

![Diagram showing a network with nodes labeled Input and Output, connecting through nodes labeled Y, Z, $X^0_{YZ}$, $X^{r-1}_{YZ}$, and $X^r_{YZ}$, with dimensions 10D, 4D, and 2D indicated.](image)
Shared Shape Space - Inference

Step 1: find a latent space point in $X_{YS}$ which produces Y that best segments the image
Shared Shape Space - Inference

Step 2: obtain a similarity function $S$
Shared Shape Space - Inference

Step 3: pick the peaks in $s$, which are latent points in space $X_{YZ}$
Shared Shape Space - Inference

Step 4: take these points as initializations, and search for latent points in $X_{YZ}$ that best segments the image.
Shared Shape Space - Inference

Step 5: obtain latent points in the upper layer and continue.
Shared Shape Space - Inference

Step 5: obtain latent points in the upper layer and continue.
Shared Shape Space - Inference

Step 5: obtain latent points in the upper layer and continue.

Please refer to the paper to see how to find the latent points for best segmenting the image in each step! (Also use GPLVM)
Shared Shape Space - Results
Tracking
Shared Shape Space - Results

Shape from single images
Shared Shape Space - Results

Gaze tracking
Outline

• A traditional generative model
• Shared shape space
• A benchmark for 3D object detection
A Benchmark: Beyond PASCAL

Benchmark for what?
A Benchmark: Beyond PASCAL

Advantage over previous benchmarks

• Background clutter
• Occluded/truncated objects
• Large number of classes and intra-class variation
• Large number of viewpoints
Advantages

• Background clutter
• Occluded/truncated objects
• Large number of classes and intra-class variation
• Large number of viewpoints
Advantages

- Background clutter
- Occluded/truncated objects
- Large number of classes and intra-class variation
- Large number of viewpoints
Advantages

- Background clutter
- Occluded/truncated objects
- Large number of classes and intra-class variation
- Large number of viewpoints
Advantages

- Background clutter
- Occluded/truncated objects
- Large number of classes and intra-class variation
- Large number of viewpoints
A Benchmark: Beyond PASCAL

Statistics of azimuth distribution
A Benchmark: Beyond PASCAL

Statistics of occlusion distribution
A Benchmark: Beyond PASCAL

Statistics of occlusion distribution
A Benchmark: Beyond PASCAL
Papers to be covered in this lecture

Tracking

• Simultaneous Monocular 2D Segmentation, 3D Pose Recovery and 3D Reconstruction [Prisacariu et al. 12]
• Monocular Multiview Object Tracking with 3D Aspect Parts [Xiang et al. 14 (ECCV)]
Tracking

Input: a sequence of frames
Output: segmentation, pose, 3D reconstruction per frame
## Outline

<table>
<thead>
<tr>
<th>Paper</th>
<th>Shape representation</th>
<th>Motion prior</th>
<th>Shape consistency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prisacariu et al. 12</td>
<td>DCT (same as Shared Shape Space)</td>
<td>None</td>
<td>Constrained to be the same 3D shape</td>
</tr>
<tr>
<td>Xiang et al. 14</td>
<td>Assembly of 3D aspect parts</td>
<td>Gaussian priors</td>
<td>Gaussian priors</td>
</tr>
<tr>
<td>Paper</td>
<td>Shape representation</td>
<td>Motion prior</td>
<td>Shape consistency</td>
</tr>
<tr>
<td>---------------------</td>
<td>-------------------------------------------</td>
<td>--------------------</td>
<td>----------------------------</td>
</tr>
<tr>
<td>Prisacariu et al. 12</td>
<td>DCT (same as Shared Shape Space)</td>
<td>None</td>
<td>Constrained to be the same 3D shape</td>
</tr>
<tr>
<td>Xiang et al. 14</td>
<td>Assembly of 3D aspect parts</td>
<td>Gaussian priors</td>
<td>Gaussian priors</td>
</tr>
</tbody>
</table>
Tracking with 3D Aspect Parts
Tracking with 3D Aspect Parts

To maximize the following posterior distribution:

\[
P(X_t, V_t | Z_{1:t}) \propto P(Z_t | X_t, V_t) \int_{X_{t-1}, V_{t-1}} P(X_t, V_t | X_{t-1}, V_{t-1}) P(X_{t-1}, V_{t-1} | Z_{1:t-1}) dX_{t-1} dV_{t-1}
\]

\[X_t\] : positions of aspect parts

\[V_t\] : viewpoints

\[Z_t\] : image
Tracking with 3D Aspect Parts

To maximize the following posterior distribution:

\[
P(X_t, V_t | Z_{1:t}) \propto P(Z_t | X_t, V_t) \int P(X_t, V_t | X_{t-1}, V_{t-1}) P(X_{t-1}, V_{t-1} | Z_{1:t-1}) dX_{t-1} dV_{t-1}
\]

- \( X_t \) : positions of aspect parts
- \( V_t \) : viewpoints
- \( Z_t \) : image
Likelihood term

Assumption of independency:

\[ P(Z_t|X_t, V_t) = \prod_{i=1}^{n} P(Z_t|X_{it}, V_t) \]

\( X_{it} \) : position of an aspect part
Likelihood term

Using classifiers in both category and instance level:

\[
P(Z_t|X_{it}, V_t) \propto \exp \left( \Lambda_{\text{category}}(Z_t, X_{it}, V_t) + \Lambda_{\text{online}}(Z_t, X_{it}, V_t) \right)
\]

\[
\Lambda_{\text{category}}(Z_t, X_{it}, V_t) = \begin{cases} 
\mathbf{w}_i^T \phi(Z_t, X_{it}, V_t), & \text{if visible} \\
\alpha_i, & \text{if self-occluded,}
\end{cases}
\]

\[
\Lambda_{\text{online}}(Z_t, X_{it}, V_t) = \begin{cases} 
\mathbf{H}_i(\psi(Z_t, X_{it}, V_t)), & \text{if visible} \\
\lambda_0, & \text{if self-occluded,}
\end{cases}
\]
Likelihood term

Descriptor of a rectified aspect part $\phi(Z_t, X_{it}, V_t)$

Rectification $\phi(Z_t, X_{it}, V_t)$

HOG

Category-level part templates $W$

Extracting Haar-like features for instance-level templates
Tracking with 3D Aspect Parts

To maximize the following posterior distribution:

\[
P(X_t, V_t | Z_{1:t}) \propto \underbrace{P(Z_t | X_t, V_t)} \int \underbrace{P(X_t, V_t | X_{t-1}, V_{t-1})} \underbrace{P(X_{t-1}, V_{t-1} | Z_{1:t-1})} dX_{t-1} dV_{t-1}
\]

\[
X_t \ : \ \text{positions of aspect parts}
\]
\[
V_t \ : \ \text{viewpoints}
\]
\[
Z_t \ : \ \text{image}
\]
Tracking with 3D Aspect Parts

To maximize the following posterior distribution:

\[
P(X_t, V_t | Z_{1:t}) \propto \frac{P(Z_t | X_t, V_t)}{P(Z_t | X_t, V_t)} \int P(X_t, V_t | X_{t-1}, V_{t-1}) P(X_{t-1}, V_{t-1} | Z_{1:t-1}) dX_{t-1} dV_{t-1}
\]

- \(X_t\) : positions of aspect parts
- \(V_t\) : viewpoints
- \(Z_t\) : image
Motion prior term

\[ P(X_t, V_t|X_{t-1}, V_{t-1}) \]
\[ = P(X_t|X_{t-1}, V_{t-1}, V_t)P(V_t|X_{t-1}, V_{t-1}) \]
\[ = P(X_t|X_{t-1}, V_t)P(V_t|V_{t-1}), \]
Motion prior term

Assumption of independency:

\[ P(X_t|X_{t-1}, V_t) \propto \prod_{i=1}^{n} P(X_{it}|X_{i(t-1)}) \prod_{(i,j)} \Lambda(X_{it}, X_{jt}, V_t), \]

\[ P(X_{it}|X_{i(t-1)}) \sim \mathcal{N}(X_{i(t-1)}, \sigma_x^2, \sigma_y^2) \]
Motion prior term

Pairwise term:

\[ \Lambda(X_{it}, X_{jt}, V_t) = P(\Delta_t(x_i, x_j)|V_t)P(\Delta_t(y_i, y_i)|V_t), \]
\[ \quad P(\Delta_t(x_i, x_j)|V_t) \sim \mathcal{N}(d_{ij,O,V_t}^x, \sigma_{dx}^2), \]
\[ \quad P(\Delta_t(y_i, y_j)|V_t) \sim \mathcal{N}(d_{ij,O,V_t}^y, \sigma_{dy}^2), \]
Inference

MCMC sampling
input: A video sequence $Z_{1:T}$, initial 3D aspect parts and viewpoint $(X_1, V_1)$
output: 3D aspect parts and viewpoints for the target in the video $(X_t, V_t)_{t=1}^T$

1. Initialize samples $(X_1^{(r)}, V_1^{(r)})_{r=1}^N$ for the first frame by sampling viewpoints and part locations according to the motion prior (6) based on $(X_1, V_1)$;

2. for $t \leftarrow 2$ to $T$ do

3. Initialize the MCMC sampler: randomly select a sample $(X_{t-1}^{(r)}, V_{t-1}^{(r)})$ as the initial state of the $(X_t, V_t)$ Markov chain;

4. repeat

5. Sample a new viewpoint from the Gaussian proposal density $Q(V_t'; V_t)$;

6. Compute the visibility of 3D aspect parts under viewpoint $V_t'$;

7. foreach part $i$ visible in both $V_t'$ and $V_t$ do

8. Sample its location from the Gaussian proposal density $Q(X_{it}'; X_{it})$;

9. end

10. foreach part $i$ visible in $V_t'$ but not in $V_t$ do

11. Compute its location $X_{it}'$ using the mean distance with respect to other visible parts according to the pairwise distributions (10);

12. end

13. Compute the acceptance ratio

$$a = \min \left(1, \frac{P(X_t', V_t'|Z_{1:t})Q(X_t; X_t')Q(V_t; V_t')}{P(X_t, V_t|Z_{1:t})Q(X_t'; X_t)Q(V_t'; V_t)} \right);$$

14. Accept the sample $(X_t', V_t')$ with probability $a$. If accepted, $(X_t, V_t) \leftarrow (X_t', V_t')$.

Otherwise, leave $(X_t, V_t)$ unchanged;

15. until $N$ samples are accepted;

16. Obtain the new sample set $(X_t^{(r)}, V_t^{(r)})_{r=1}^N$, and find the MAP among it as the tracking output for frame $t$;

end
## Results - Tracking

<table>
<thead>
<tr>
<th>Video</th>
<th>MIL 2</th>
<th>L1 3</th>
<th>TLD 20</th>
<th>Struct 15</th>
<th>DPM 12+PF</th>
<th>Category Model</th>
<th>Full Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Race1</td>
<td>0.34</td>
<td>0.39</td>
<td>0.20</td>
<td>0.36</td>
<td>0.68</td>
<td>0.68</td>
<td>0.69</td>
</tr>
<tr>
<td>Race2</td>
<td>0.49</td>
<td>0.49</td>
<td>0.28</td>
<td>0.50</td>
<td>0.74</td>
<td>0.74</td>
<td>0.73</td>
</tr>
<tr>
<td>Race3</td>
<td>0.36</td>
<td>0.26</td>
<td>0.25</td>
<td>0.44</td>
<td>0.74</td>
<td>0.74</td>
<td>0.77</td>
</tr>
<tr>
<td>Race4</td>
<td>0.53</td>
<td>0.56</td>
<td>0.47</td>
<td>0.63</td>
<td>0.76</td>
<td>0.76</td>
<td>0.76</td>
</tr>
<tr>
<td>Race5</td>
<td>0.29</td>
<td>0.54</td>
<td>0.28</td>
<td>0.26</td>
<td>0.63</td>
<td>0.63</td>
<td>0.68</td>
</tr>
<tr>
<td>Race6</td>
<td>0.27</td>
<td>0.53</td>
<td>0.48</td>
<td>0.29</td>
<td>0.76</td>
<td>0.76</td>
<td>0.77</td>
</tr>
<tr>
<td>SUV1</td>
<td>0.58</td>
<td>0.81</td>
<td>0.56</td>
<td>0.60</td>
<td>0.78</td>
<td>0.78</td>
<td>0.78</td>
</tr>
<tr>
<td>SUV2</td>
<td>0.18</td>
<td>0.12</td>
<td>0.53</td>
<td>0.24</td>
<td>0.77</td>
<td>0.77</td>
<td>0.77</td>
</tr>
<tr>
<td>Sedan</td>
<td>0.26</td>
<td>0.23</td>
<td>0.33</td>
<td>0.30</td>
<td>0.78</td>
<td>0.78</td>
<td>0.78</td>
</tr>
<tr>
<td><strong>Mean</strong></td>
<td>0.37</td>
<td>0.44</td>
<td>0.38</td>
<td>0.40</td>
<td>0.74</td>
<td>0.74</td>
<td>0.75</td>
</tr>
<tr>
<td>KITTI01</td>
<td>0.20</td>
<td>0.40</td>
<td>0.44</td>
<td>0.33</td>
<td>0.65</td>
<td>0.65</td>
<td>0.64</td>
</tr>
<tr>
<td>KITTI02</td>
<td>0.28</td>
<td>0.18</td>
<td>0.20</td>
<td>0.12</td>
<td>0.26</td>
<td>0.26</td>
<td>0.32</td>
</tr>
<tr>
<td>KITTI03</td>
<td>0.37</td>
<td>0.59</td>
<td>0.42</td>
<td>0.36</td>
<td>0.20</td>
<td>0.20</td>
<td>0.50</td>
</tr>
<tr>
<td>KITTI04</td>
<td>0.31</td>
<td>0.12</td>
<td>0.36</td>
<td>0.34</td>
<td>0.67</td>
<td>0.67</td>
<td>0.33</td>
</tr>
<tr>
<td>KITTI05</td>
<td>0.40</td>
<td>0.32</td>
<td>0.51</td>
<td>0.41</td>
<td>0.54</td>
<td>0.54</td>
<td>0.72</td>
</tr>
<tr>
<td>KITTI06</td>
<td>0.64</td>
<td>0.21</td>
<td>0.54</td>
<td>0.65</td>
<td>0.65</td>
<td>0.65</td>
<td>0.56</td>
</tr>
<tr>
<td>KITTI07</td>
<td>0.12</td>
<td>0.33</td>
<td>0.03</td>
<td>0.28</td>
<td>0.66</td>
<td>0.66</td>
<td>0.66</td>
</tr>
<tr>
<td>KITTI08</td>
<td>0.58</td>
<td>0.13</td>
<td>0.00</td>
<td>0.66</td>
<td>0.74</td>
<td>0.74</td>
<td>0.72</td>
</tr>
<tr>
<td>KITTI09</td>
<td>0.18</td>
<td>0.15</td>
<td>0.17</td>
<td>0.18</td>
<td>0.51</td>
<td>0.51</td>
<td>0.52</td>
</tr>
<tr>
<td>KITTI10</td>
<td>0.33</td>
<td>0.46</td>
<td>0.41</td>
<td>0.35</td>
<td>0.68</td>
<td>0.68</td>
<td>0.68</td>
</tr>
<tr>
<td>KITTI11</td>
<td>0.28</td>
<td>0.23</td>
<td>0.24</td>
<td>0.28</td>
<td>0.71</td>
<td>0.71</td>
<td>0.68</td>
</tr>
<tr>
<td><strong>Mean</strong></td>
<td>0.34</td>
<td>0.28</td>
<td>0.29</td>
<td>0.36</td>
<td>0.54</td>
<td>0.54</td>
<td>0.58</td>
</tr>
<tr>
<td>06_car 20</td>
<td>0.19</td>
<td>0.52</td>
<td>0.85</td>
<td>0.48</td>
<td>0.70</td>
<td>0.67</td>
<td>0.70</td>
</tr>
</tbody>
</table>
Results – Viewpoint and 3D recovery

<table>
<thead>
<tr>
<th>Video</th>
<th>Viewpoint Estimation</th>
<th>3D Aspect Part Localization</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Full Model</td>
<td>Category Model</td>
</tr>
<tr>
<td>Race1</td>
<td>0.67/18.73°</td>
<td>0.59/22.88°</td>
</tr>
<tr>
<td>Race2</td>
<td>0.77/10.83°</td>
<td>0.60/12.65°</td>
</tr>
<tr>
<td>Race3</td>
<td>0.83/9.28°</td>
<td>0.83/7.79°</td>
</tr>
<tr>
<td>Race4</td>
<td>0.69/15.83°</td>
<td>0.68/14.67°</td>
</tr>
<tr>
<td>Race5</td>
<td>0.71/10.75°</td>
<td>0.74/11.78°</td>
</tr>
<tr>
<td>Race6</td>
<td>0.43/18.47°</td>
<td>0.40/21.34°</td>
</tr>
<tr>
<td>SUV1</td>
<td>0.82/7.81°</td>
<td>0.75/8.52°</td>
</tr>
<tr>
<td>SUV2</td>
<td>0.57/19.56°</td>
<td>0.45/56.33°</td>
</tr>
<tr>
<td>Sedan</td>
<td>0.76/9.87°</td>
<td>0.78/9.50°</td>
</tr>
<tr>
<td>Mean</td>
<td>0.69/13.46°</td>
<td>0.65/18.38°</td>
</tr>
<tr>
<td>KITTI01</td>
<td>0.95/6.54°</td>
<td>0.74/8.53°</td>
</tr>
<tr>
<td>KITTI02</td>
<td>1.00/5.40°</td>
<td>0.20/30.06°</td>
</tr>
<tr>
<td>KITTI03</td>
<td>0.42/15.64°</td>
<td>0.42/15.14°</td>
</tr>
<tr>
<td>KITTI04</td>
<td>0.22/27.05°</td>
<td>0.25/26.03°</td>
</tr>
<tr>
<td>KITTI05</td>
<td>0.36/23.59°</td>
<td>0.40/22.17°</td>
</tr>
<tr>
<td>KITTI06</td>
<td>0.31/21.63°</td>
<td>0.29/21.58°</td>
</tr>
<tr>
<td>KITTI07</td>
<td>0.96/6.86°</td>
<td>0.89/7.92°</td>
</tr>
<tr>
<td>KITTI08</td>
<td>0.57/15.61°</td>
<td>0.48/23.84°</td>
</tr>
<tr>
<td>KITTI09</td>
<td>0.50/21.63°</td>
<td>0.42/78.67°</td>
</tr>
<tr>
<td>KITTI10</td>
<td>0.81/7.99°</td>
<td>0.79/9.44°</td>
</tr>
<tr>
<td>KITTI11</td>
<td>0.88/9.33°</td>
<td>0.78/11.80°</td>
</tr>
<tr>
<td>Mean</td>
<td>0.63/14.66°</td>
<td>0.51/23.20°</td>
</tr>
</tbody>
</table>
Papers to be covered in this lecture

Shape synthesis

• A Probabilistic Model for Component-based Shape Synthesis [Kalogerakis et al. 12]
• Co-constrained Handles for Deformation in Shape Collections [Yumer et al. 14]
Papers to be covered in this lecture

Shape synthesis

- A Probabilistic Model for Component-based Shape Synthesis [Kalogerakis et al. 12]
- Co-constrained Handles for Deformation in Shape Collections [Yumer et al. 14]
Probabilistic Model for Shape Synthesis

Input

Probabilistic model

Learning stage

Synthesis stage

Output

(from Kalogerakis’ SIGGRAPH slides)
Learning stage

Input

Probabilistic model

Learning stage
$P(R) \prod_{l \in L} \left[ P(\mathcal{N}_l | R) \right]$
\[ P(R) \prod_{l \in L} \left[ P(N_l | R) P(S_l | R) \right] \]
\[ P(R) \prod_{l \in L} \left[ P(N_l | R) P(S_l | R) \right] \]
\[ P(R) \prod_{l \in L} \left[ P(N_l | R) P(S_l | R) P(D_l | S_l) P(C_l | S_l) \right] \]
\[ P(R) \prod_{l \in L} \left[ P(N_l | R) P(S_l | R) P(D_l | S_l) P(C_l | S_l) \right] \]
Let's consider a hierarchical model with latent variables. The diagram illustrates the relationships between the latent objects and components, with each node representing a variable:

- **N_l** (Latent object) and **S_l** (Latent component) are connected in the hierarchy.
- **D_l** and **C_l** are components of **S_l**.

The model can be mathematically represented as:

\[
P(R) \prod_{l \in L} \left[ P(N_l | R) P(S_l | R) P(D_l | S_l) P(C_l | S_l) \right]
\]
Learn from training data:
- latent styles
- lateral edges
- parameters of CPDs
Learning stage

Given observed data $\mathbf{O}$, find structure $\mathbf{G}$ that maximizes:

$$P(G \mid \mathbf{O}) = \frac{P(\mathbf{O} \mid G)P(G)}{P(\mathbf{O})}$$
Learning stage

Given observed data \( \mathbf{O} \), find structure \( \mathbf{G} \) that maximizes:

\[
P(G \mid \mathbf{O}) = \frac{P(\mathbf{O} \mid G)P(G)}{P(\mathbf{O})}
\]

Assuming uniform prior over structures, maximize marginal likelihood:

\[
P(\mathbf{O} \mid G) = \sum_{R,S} \int P(\mathbf{O}, R, S \mid \Theta, G)P(\Theta \mid G) \, d\Theta
\]
Learning stage

Given observed data $\mathbf{O}$, find structure $\mathbf{G}$ that maximizes:

$$P(\mathbf{G} | \mathbf{O}) = \frac{P(\mathbf{O} | \mathbf{G})P(\mathbf{G})}{P(\mathbf{O})}$$

Assuming uniform prior over structures, maximize marginal likelihood:

$$P(\mathbf{O} | \mathbf{G}) \approx P(\mathbf{O^*} | \mathbf{G}) \cdot \frac{P(\mathbf{O} | \mathbf{G}, \tilde{\Theta}_G)}{P(\mathbf{O^*} | \mathbf{G}, \tilde{\Theta}_G)}$$

*(Cheeseman-Stutz approximation)*
Synthesis Stage

Probabilistic model

Output

(from Kalogerakis’ SIGGRAPH slides)
Synthesizing a set of components

Enumerate high-probability instantiations of the model

\[
\begin{align*}
R & \rightarrow N \rightarrow S \\
C & \rightarrow D \\
\{R=1\} & \rightarrow \{R=1, S_1=1\}, \{R=1, S_1=2\} \\
\{R=2\} & \rightarrow \{R=2, S_1=2\}, \{R=2, S_1=2\} \\
\end{align*}
\]

...
Optimizing component placement

Source shapes

Unoptimized new shape

Optimized new shape
Results

New shape

Source shapes (colored parts are selected for the new shape)
Results of alternative models: no latent variables
Results of alternative models: no part correlations
Papers to be covered in this lecture

Shape synthesis

• A Probabilistic Model for Component-based Shape Synthesis [Kalogerakis et al. 12]

• Co-constrained Handles for Deformation in Shape Collections [Yumer et al. 14]
Co-Constrained Handles for Shape Deformation

Input

Constraints

Learning stage

Output

Synthesis stage
Pipeline

Initial Segmentation

Initial Abstraction

Clustering

Co-constrained Abstraction

Handles

Constraints
Pipeline

Initial Segmentation

Initial Abstraction

Clustering

Co-constrained Abstraction

Handles

Constraints
Clustering

- Seed surface clustering
- Global Clustering
Clustering

• Seed surface clustering
  – Aligning semantically similar/identical segments

• Global Clustering
  – Creating larger clusters that join the surfaces from different segments
  – Based on geometric features of the surfaces, including PCA basis, normal and curvature signature of the surface.
Clustering

Seed clustering

Door handle

Wheels

Different clusters
Clustering

Global clustering

Door handle
Wheels
Same cluster
Pipeline

Initial Segmentation

Initial Abstraction

Clustering

Co-constrained Abstraction

Handles

Constraints
Co-constrained abstraction

Planer, spherical, or cylindrical?

Initial abstraction

Co-abstraction
Identify constraints of DOFs

Identify as constraints if variance is below certain threshold

• Translation
• Rotation
• Scaling
Pipeline

Initial Segmentation

Initial Abstraction

Clustering

Co-constrained Abstraction

Handles

Constraints
Correlation among deformation handles

Estimate the conditional feature probability

\[ P(k_i | k_j) = \frac{P(k_i, k_j)}{P(k_j)} \]

where \( k \) takes values from: \( t_x, t_y, t_z, \theta_x, \theta_y, \theta_z, s_x, s_y, s_z, r, l \).
Results
Summary

3D shape from image or video

• Using silhouettes, colors or features (e.g. HOG) in image
• Using voxel occupancies, aspect parts, level set, or meshes to represent 3D shapes
• Either generative model, shared shape space, or discriminative model
Summary

Shape synthesis

• Generative model based on semantic parts, which models styles also.
• Modeling correlation between geometric features.