Face detection, pose estimation and landmark localization in the wild

Presenter: Shuai Zheng (Kyle)
Paper: X. Zhu and D. Ramanan in CVPR 2012
Many Applications of Face Det, Pose Est, Landmarks Loc.

- Camera auto focus
- Animation
- Organize photos
- Face recognition
- Expression analysis
- Human interaction
Hot Area

**Microsoft** Face Game

**Google** Picasa’s Face Movie

Face.com App (**Facebook**)

**Microsoft® Research**

Face Apps

and Facial expression recognition, etc…
Face Recognition Pipeline

How far is our technique from 100% accuracy face recognition (object recognition) system?

Assume the previous step is perfect. Overly optimistic!

Name: Andrea?
Gender: Male
Age: 24?
Has beard?

Face recognition in the wild

- Face presents different appearances and shapes under different viewpoints;
Face presents different appearances and shapes under different elastic deformation.
Optimizing all isolated components in a computer vision system is very difficult.

Viewpoints problem

Elastic deformation problem

Do we need to collect billions of low-quality data to get state-of-the-art?
Structured SVM with mixtures of trees
Model viewpoints with mixtures of trees

- Model topological changes with view-specific mixtures (simple way to capture 3D!)

Ioffe & Forsyth, 2001
Everingham, Sivic, & Zisserman, 2006
Model viewpoints with mixtures of trees

- Model topological changes with view-specific mixtures (simple way to capture 3D!)
- Branches appear and disappear.
- Capture self-occlusion due to viewpoint changes
Model elastic deformations with trees

- Tree-structure elastic model handles deformation
- Efficient optimization with dynamic programming

Felzenszwalb & Huttenlocher 2005
Pictorial Structured Model

$I$: Image \quad $L$: locations of parts \quad $m$: mixture
Inference

For all locations:

$$(m^*, L^*) = \arg \max_m \left( \max_L S(I, L, m) \right)$$

For a tree $(V,E)$: dynamic programing

$m^*$ : the estimated viewpoint.
$L^*$ : the estimated landmark locations.
Search over scales using an image pyramid.
Learning

• Fully supervised data

• Learn the trees using Chow-Liu algorithm
  Chow & Liu, 1968

• Learn the appearance and deformation jointly using a structured SVM

\[ S(I, L, m) = \beta \cdot \Phi(I, L, m) \]
Chow-Liu algorithm is an efficient method for constructing a second-order product approximation of a joint distribution.

Joint probability distribution $P(X_1, \ldots, X_n)$ can be described as a product of second-order conditional and marginal distributions. As shown in the figure,

$$P'(X_1, X_2, X_3, X_4, X_5, X_6) = P(X_6|X_5)P(X_5|X_2)P(X_4|X_2)P(X_3|X_2)P(X_2|X_1)P(X_1)$$
Learning with structured SVM

- Learn the appearance and deformation jointly using a structured SVM

\[ S(I, L, m) = \beta \cdot \Phi(I, L, m) \]

\[ S(I, L, m) = \text{App}_m(I, L) + \text{Shape}_m(L) + \alpha^m \]  \hspace{1cm} (1)

\[ \text{App}_m(I, L) = \sum_{i \in V_m} w_i^m \cdot \phi(I, l_i) \]  \hspace{1cm} (2)

\[ \text{Shape}_m(L) = \sum_{ij \in E_m} a_{ij}^m dx^2 + b_{ij}^m dx + c_{ij}^m dy^2 + d_{ij}^m dy \]  \hspace{1cm} (3)
Problem Formulation

Given labeled positive examples \( \{I_n, L_n, m_n\} \) and negative examples \( \{I_n\} \), let's write \( z_n = \{L_n, m_n\} \). The score function is linear in the part templates \( w \), spring parameters \( (a, b, c, d) \) and mixture biases \( \alpha \). Concatenated all the parameters into \( \beta \). We can formulate the problem as

\[
\arg \min_{\beta, \xi_n \geq 0} \frac{1}{2} \beta \cdot \beta + C \sum_n \xi_n
\]

s.t. \( \forall n \in \text{pos} \) \( \beta \cdot \Phi(I_n, z_n) \geq 1 - \xi_n \)
\( \forall n \in \text{neg}, \forall z \) \( \beta \cdot \Phi(I_n, z) \leq -1 + \xi_n \)
\( \forall k \in K, \beta_k \leq 0 \)
Experimental Results

Detection

Winner of LFW face recognition challenge
Experimental Results

Detection

Dalal & Triggs, 2005

Multiview

- Multiview HoG
- Google Picasa
- face.com
- 2-view Viola Jones
Experimental Results

Detection

Felzenszwalb et al. 2009

Def. Parts

Multiview

DPM, voc-re4
Multiview HoG
Google Picasa
face.com
2-view Viola Jones
Experimental Results

Detection

- Our star model
- DPM, voc-re4
- Multiview HoG
- Google Picasa
- face.com
- 2-view Viola Jones

Precision vs. Recall Chart:
- Supv. Parts
- Def. Parts
- Multiview

Recall:
- 0.1
- 0.2
- 0.3
- 0.4
- 0.5
- 0.6
- 0.7
- 0.8
- 0.9

Precision:
- 0.2
- 0.3
- 0.4
- 0.5
- 0.6
- 0.7
- 0.8
- 0.9

Inset Image:
- Star model with red and black segments.
Experimental Results

Detection

- Our model
- Our star model
- DPM, voc-re4
- Multiview HoG
- Google Picasa
- face.com
- 2-view Viola Jones

Tree structure

Supv. Parts & Mixtures

Def. Parts

Multiview
Experimental Results

Detection

Our model is learned with only **hundreds** of faces as opposed to billions!

### Experimental Results

<table>
<thead>
<tr>
<th>Pose</th>
<th>Landmark</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Multiview HoG</td>
<td>• Multiview AAMs (Cootes et al. 2000)</td>
</tr>
<tr>
<td>• Multiview AAMs</td>
<td>• Oxford (Everingham et al. 2006)</td>
</tr>
<tr>
<td>• face.com</td>
<td>• CLM (Saragih et al. 2011)</td>
</tr>
</tbody>
</table>

**Give the baselines unfair advantages by:**
- Initialized with *ground truth* detection.

**Harsh on our model:**
- Count the missed faces as having inf. pose and localization error.
Experimental Results

Pose estimation

MultiPIE

AFW

Multiview AAMs are initialized with ground truth.
Experimental Results

Landmark localization

MultiPIE

AFW

Multiview AAM, CLM, Oxford are initialized with ground truth.
Conclusions

Pros:

• Model the view-specific within mixtures of trees.
• Joint method to do face detection, pose estimation, and landmarks localization for face images with viewpoint variations and elastic deformation.
Conclusions

Cons:

- Slow in the inference, given one image (80*80), it takes more than 20 seconds to process.
- Cannot handle large size images.
Conclusions

Messages:

- Tree-structure elastic model can do many jobs together.
- Matching small patch is much easier than matching the object of interest.
- Training model on selective supervised data is the key to success.
Thank you!