Simultaneous Cascaded Regression

Pedro Martins, Jorge Batista
http://www.isr.uc.pt/~pedromartins
pedromartins@isr.uc.pt

Institute of Systems and Robotics (ISR)
University of Coimbra
Portugal
Introduction

• Facial landmark localization with deformable models (nonrigid face alignment)

• Lucas & Kanade Image Alignment Framework
  - Simultaneous Forwards Additive / Inverse Compositional Algorithm

• Cascaded Regression Framework

• Simultaneous Algorithm: Cascaded Regression Extension
  - Regression w/ both shape and appearance structure
Outline

• Related Work

• Base Components
  - Warp Function
  - Parametric Models of Shape and Appearance

• Lucas & Kanade Image Alignment Framework
  - Simultaneous Forwards Additive (SFA)
  - Simultaneous Inverse Compositional (SIC)

• Simultaneous Cascaded Regression (SCR)

• Evaluation Results (LFPW, HELEN, LFW, 300W datasets)
Related Work

- Active Shape Model (ASM)
- Deformable Part Model (DPM)
- Active Appearance Model (AAM)
  - Project-Out Inverse Compositional (PO-IC)
  - Simultaneous Inverse Compositional (SIC)
- Constrained Local Model (CLM)
  - Convex Quadratic Fitting (CQF)
  - Subspace Constrained Mean-Shifts (SCMS)
  - Bayesian CLM (BCLM)
- Cascaded Regression (CR)
  - Supervised Descent Method (SDM)
  - Project-Out Cascade Regression (PO-CR)
Newton Methods vs Cascaded Regression

Newton's Method Optimization

Cost Function
Piecewise Affine Warp (@AAMs) [Not Used Here]
Patch based Local Warp

Landmarks

Similarity Warp \((s, \theta, t_x, t_y)\)

Local Patches

Local Extracted Features (HoG)
Parametric Shape and Appearance Models

\[
\mathbf{s} = (x_1, \ldots, x_v, y_1, \ldots, y_v)^T \in \mathbb{R}^{2v}
\]

Shape Model

\[
\mathcal{W}(\mathbf{s}; \mathbf{p}) = \mathbf{s}_0 + \sum_{i=1}^{n+4} \phi_i p_i
\]

Appearance Model

\[
A(x; \lambda) = A_0(x) + \sum_{i=1}^{m} A_i(x) \lambda_i
\]
Local Appearance Model (LAM)

• Combined Parametric Model
  • Shape Regularization
  • Local Appearance (w/ HoG Features)

• Model Optimization/Fitting
  • Linear Warp Function
  • LK Framework
  • Cascaded Regression

$\mathcal{M}(p, \lambda) \equiv \mathcal{W}(s; p) \bigcup \mathcal{A}(x; \lambda)$

Shape + Pose
Appearance
Simultaneous Forwards Additive (SFA)

Goal
\[
\arg \min_{p, \lambda} \|A_0 + A\lambda - I(\mathcal{W}(p))\|^2
\]

Iteratively solve for small updates
\[
\arg \min_{\Delta p, \Delta \lambda} \|A_0 + A(\lambda + \Delta \lambda) - I(\mathcal{W}(p + \Delta p))\|^2
\]

Solution
\[
\begin{bmatrix}
\Delta p \\
\Delta \lambda
\end{bmatrix} = H^{-1}_{FA} J^T_{FA} \left[ A_0 + A\lambda - I(\mathcal{W}(p)) \right]
\]

Jacobian
\[
J_{FA} = \left( \nabla I \frac{\partial \mathcal{W}(p)}{\partial p}, A \right)
\]

Hessian
\[
H_{FA} = J^T_{FA} J_{FA}
\]

Parameters Update
\[
p \leftarrow p + \Delta p \\
\lambda \leftarrow \lambda + \Delta \lambda
\]
Simultaneous Inverse Compositional (SIC)

Goal
\[
\arg \min_{p, \lambda} \| A_0 + A\lambda - I(W(p)) \|^2
\]

Iteratively solve for small updates
\[
\arg \min_{\Delta p, \Delta \lambda} \| A_0(W(\Delta p)) + A(W(\Delta p))(\lambda + \Delta \lambda) - I(W(p)) \|^2
\]

Solution
\[
\begin{bmatrix}
\Delta p \\
\Delta \lambda
\end{bmatrix} = -H_{IC}^{-1}J_{IC}^T [A_0 + A\lambda - I(W(p))]
\]

Jacobian
\[
J_{IC} = \begin{pmatrix}
\nabla A_0 + \nabla A\lambda \\
\frac{\partial W(0)}{\partial p}, A
\end{pmatrix}
\]

Hessian
\[
H_{IC} = J_{IC}^T J_{IC}
\]

Parameters Update
\[
W(s, p) \leftarrow W(s, p) \circ W(s, \Delta p)^{-1} \\
p \leftarrow p - \Delta p \\
\lambda \leftarrow \lambda + \Delta \lambda
\]

Gauss Newton Approximation
Simultaneous Cascaded Regression (SCR)

Regression with both shape and appearance structure

\[
\begin{bmatrix}
  p \\
  \lambda
\end{bmatrix}^k = \begin{bmatrix}
  p \\
  \lambda
\end{bmatrix}^{k-1} + R^{k-1} \left( \mathbf{W}(p^{k-1}) - A_0 - A\lambda^{k-1} \right), \quad k = 1, \ldots, K
\]

Features extracted at previous level
Features generated by the Model

Simultaneous Inverse Compositional (SIC) vs Simultaneous Cascaded Regression (SCR)

Cost Function
SCR - Learning Regression Matrices

Estimate average Jacobian under multiple initializations

\[
\arg \min_{J^k_S} \sum_{i=1}^N \int p(r') \left\| A_0 + A_\lambda_i^k + J^k_S \Delta r_i^k - I_i(\mathcal{W}(p_i^k)) \right\|^2 dr'
\]

Deviation from Ground Truth

\[
\Delta r_i^k = \begin{bmatrix}
p_i^k - p_* \\
\lambda_i^k - \lambda_*
\end{bmatrix}
\]

Discrete approximation

\[
\arg \min_{J^k_S} \sum_{i=1}^N \sum_{j=1}^M \left\| A_0 + A_\lambda_{ij}^k + J^k_S \Delta r_{ij}^k - I_i(\mathcal{W}(p_{ij}^k)) \right\|^2
\]

Solution by Ridge Regression

\[
J^k_S = \left( \Delta r \Delta r^T + \lambda_1 I_d \right)^{-1} \Delta r \ E^T
\]

Advantage: do not require to invert a large data matrix

Update matrix

\[
R^k = \left( (J^k_S)^T J^k_S + \lambda_2 I_d \right)^{-1} (J^k_S)^T
\]

Cascade update

\[
\Delta r^k = R^k \left( I(\mathcal{W}(p^k)) - A_0 - A_\lambda^k \right)
\]

\[
r^{k+1} = r^k + \Delta r^k
\]

E: Data matrix w/ entries

\[
E_{ij} = I_i(\mathcal{W}(p_{ij}^k)) - A_0 - A_\lambda_{ij}^k
\]
Cascaded Regression Learning
## Evaluation Results

### Cumulative error distribution function (CDF)

<table>
<thead>
<tr>
<th>Method / AUC</th>
<th>LFPW</th>
<th>HELEN</th>
<th>LFW</th>
<th>300W</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial Estimate</td>
<td>46.4</td>
<td>41.6</td>
<td>61.7</td>
<td>27.2</td>
</tr>
<tr>
<td>PO-FA</td>
<td>53.6</td>
<td>51.3</td>
<td>67.3</td>
<td>38.2</td>
</tr>
<tr>
<td>SFA</td>
<td>70.0</td>
<td>60.2</td>
<td>73.0</td>
<td>42.3</td>
</tr>
<tr>
<td>PO-IC</td>
<td>56.1</td>
<td>53.8</td>
<td>69.4</td>
<td>39.1</td>
</tr>
<tr>
<td>SIC</td>
<td>73.1</td>
<td>63.5</td>
<td>75.6</td>
<td>43.9</td>
</tr>
<tr>
<td>SCMS</td>
<td>56.9</td>
<td>50.7</td>
<td>70.7</td>
<td>40.9</td>
</tr>
<tr>
<td>TM</td>
<td>56.5</td>
<td>54.8</td>
<td>60.1</td>
<td>36.7</td>
</tr>
<tr>
<td>SDM</td>
<td>72.2</td>
<td>69.7</td>
<td>81.5</td>
<td>50.3</td>
</tr>
<tr>
<td>PO-CR</td>
<td>80.4</td>
<td>72.5</td>
<td>84.1</td>
<td>53.3</td>
</tr>
<tr>
<td>SCR</td>
<td><strong>82.6</strong></td>
<td><strong>74.8</strong></td>
<td><strong>85.5</strong></td>
<td><strong>55.5</strong></td>
</tr>
</tbody>
</table>

**Inter-ocular normalized error**

\[
e_m(s) = \frac{1}{v} \sum_{i=1}^{v} ||s_i - s^*_i||
\]

Area Under Curve (AUC)
Landmark Fitting Error Standard Deviation

LFPW Database

HELEN Database

300W Database

LFW Database
Qualitative Results (LFPW Database)
Qualitative Results (HELEN Database)
SCR Fitting Video

Simultaneous Cascaded Regression
Pedro Martins, Jorge Batista
ISR - Institute of Systems and Robotics
University of Coimbra, Portugal

IEEE International Conference on Image Processing 2018
ICIP 2018
Conclusions

- Facial landmark localization w/ deformable face model
- Simultaneous Algorithm: Cascaded Regression Extension
  - Regression w/ both shape and appearance structure
  - Learning stage w/o inverting a large data matrix
- Evaluation Results (LFPW, HELEN, LFW, 300W datasets)
- Acknowledgements
  - Work supported by the Portuguese Science Foundation (‘Fundação para a Ciência e Tecnologia - FCT’) through the grant SFRH/BDP/90200/2012.
Questions?

http://www.isr.uc.pt/~pedromartins
pedromartins@isr.uc.pt
SCR Fitting Video

Simultaneous Cascaded Regression
Pedro Martins, Jorge Batista
ISR - Institute of Systems and Robotics
University of Coimbra, Portugal

IEEE International Conference on Image Processing 2018
ICIP 2018