From Understanding to Controlling of Privacy against Automatic Person Identification  
Seong Joon Oh

Motivation

• How much private information is contained in social media photos.
• Tool for controlling the amount of private content in photos.

Understanding

Person recognition in personal photos.

<table>
<thead>
<tr>
<th>Task</th>
<th>Pose</th>
<th>Time</th>
<th>Input</th>
</tr>
</thead>
<tbody>
<tr>
<td>Face recognition</td>
<td>Frontal</td>
<td>Years</td>
<td>Head</td>
</tr>
<tr>
<td>Re-identification</td>
<td>Upright</td>
<td>Hours</td>
<td>Full body</td>
</tr>
</tbody>
</table>

• Less studied task in computer vision.
• Relevant for privacy in social media photos.

Method: simple yet effective [1]

• Extract convnet features from five fixed regions & concat.
• State of the art performance without relying on pose normalisation and face recogniser.

Dataset: PIPA [4]

• First social-media scale person recognition dataset.
• We introduced more challenging setups [1].

Results

<table>
<thead>
<tr>
<th>Method</th>
<th>#Cues</th>
<th>Acc %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chance</td>
<td>-</td>
<td>0.17</td>
</tr>
<tr>
<td>PIPER [4]</td>
<td>109</td>
<td>83.1</td>
</tr>
<tr>
<td>OURS</td>
<td>17</td>
<td>86.8</td>
</tr>
</tbody>
</table>

• Simple method leads to effective identification.
• There is ample identity information in social media photos that can be automatically mined.

Controlling

Usual obfuscation: unpleasant & ineffective.

• Faceless recognition techniques [2]:
  1. fine-tune features for obfuscation.
  2. graphical inference within albums.
• They re-enable recognition (12x random chance even in the worst case).

Adversarial image perturbations (AIP)

• Unlike blurring, AIPs generate visually pleasant and effective obfuscation patterns.
• Caveat: there exist defense strategies.

Game theoretical framework [3]

• Model uncertainty in the choice of attack-defense strategies as a ZSG.
• Derive privacy guarantees from explicit assumptions.

A case study

• Defense strategy: translate, noise, blur, crop.
• Attack strategy: train AIP against each defense strategy.
• Game theoretical privacy guarantee: at most 7.3%.

References