Person Re-identification

Introduction and Future Trends

Shengcai Liao
Institute of Automation
Chinese Academy of Sciences

ECCV 2018 Tutorial • Munich
Representation Learning for Pedestrian Re-identification - Schedule

- 09:00 – 09:40 Introduction and future trends, Shengcai Liao
- 09:40 – 10:20 Visual descriptors and similarity metrics, Yang Yang
- 10:20 – 10:40 Coffee break
- 10:40 – 11:40 Deep learning and transfer learning, Zhun Zhong
- 11:40 – 12:00 Questions & Discussions
PART ONE

Introduction
• Security concerns

2011 riot in London

2013 Boston Marathon bombings

2012 “8.10” serial killer Zhou Kehua

2014 “3.1” Kunming terror attack
Background

- Surveillance cameras everywhere
- However,
  - Mostly, searching suspects still requires large amount of labors
  - Automatic algorithms are still poor
  - But the real demand is increasing
Search suspects in a large amount of videos
Concepts

**Classification:** classes fixed

**Verification:** pairwise

**Identification:** gallery IDs known

**Re-identification:** gallery IDs unknown
History

Difference with Multi-camera Tracking

• Multi-camera tracking
  • Usually online
  • Need to track all persons in all cameras
  • In a local area
  • In a short duration

• Person Re-identification
  • Usually offline, for retrieval
  • Re-identify one specific person
  • Across broad areas
  • With a possible long time

Oriented from multi-camera tracking, but is a particular independent task now.
Popularity


CVPR 2018: 27
ECCV 2018: 12
Pipeline

Preprocess
- Pedestrian detection
- Single-camera Tracking

Representation
- Hand-crafted features
- Feature learning

Matching
- Traditional Distances
- Metric learning
- Re-ranking
Challenges

- Viewpoint changes
- Pose changes
- Illumination variations
- Occlusions
- Low resolutions
- Limited labeled data
- Generalization ability
Main research directions in person re-identification
Feature Design

**Color**
- RGB, HSV, YCbCr, Lab, Color names

**Texture**
- Gabor, LBP, SILTP, Schmid, BiCov

**Hybrid**
- ELF, LOMO, GOG

**Structure**
- Pictorial, SDALF, Saliency

**Attribute**
- Age, gender, bag
Feature Design

• Typical feature: LOMO
  • Illumination variations: retinex and SILTP
  • Viewpoint changes: local maximal occurrence

S. Liao et al., "Person Re-identification by Local Maximal Occurrence Representation and Metric Learning," In CVPR 2015.
Metric Learning

---

**Traditional Methods**
- ITML, LMNN, LDML

**Optimization Methods**
- PRDC, MLAPG

**Fast Methods**
- KISSME, XQDA, LSSL

---

$$D^2_M(x, z) = \|x - z\|_M^2 = (x - z)^T M (x - z)$$
Deep Learning

• Deep metric learning
  • Cosine similarity
  • Contrastive loss
  • Triplet loss
  • Center loss
Deep Learning

- Deep structures
  - Siamese CNN
  - Cross-input neighborhood, patch summary
  - Gating CNN
  - Contextual LSTM
  - Attention network
Deep Learning

- Sample mining
  - Hard negative mining
  - Moderate positive sample mining

Re-ranking

- User feedback based methods (human in the loop)
  - POP
  - HVIL
Re-ranking

- Context based methods
  - DCIA
  - Bidirectional ranking
  - DSAR

Transfer Learning

- Cross-dataset evaluation
  - Dong Yi et al. 2014, deep metric learning: cross-dataset evaluation
  - Yang Hu et al. 2014, "Cross dataset person re-identification"

- Transfer learning / domain adaptation
  - Supervised
    - Pre-train + fine tuning
  - Unsupervised
    - UMDL, CVPR 2016
    - CAMEL, ICCV 2017
    - SPGAN, CVPR 2018
    - HHL, ECCV 2018
PART THREE

Evaluation and Benchmark
Evaluation

• Closed-set scenario
  • Probe:
    • query images to be re-identified
  • Gallery:
    • a set of images from surveillance videos to re-identify probe images
• Performance measure:
  • Cumulative Matching Characteristic (CMC) curves
  • mAP: mean average precision

mAP is from image retrieval. CMC is more practical for person re-id, because one correct retrieval is already enough for forensic search.

Constraint: each probe image must have the same person appearing in the gallery
Evaluation

- Open-set scenario
Open-set Person Re-identification

- Task: determine the same person of the probe in the gallery, or reject the probe
- Two subsets of probes

- Genuine Probe $P_G$: Need to accept and re-identify, but large intra-class variations
- Impostor Probe $P_N$: Need to reject, but can be similar, e.g. similar frontal view
Open-set Person Re-identification

• Performance measures:
  • Detection and Identification Rate (DIR): percentage of images in $P_G$ that are correctly accepted and re-identified
  • False Accept Rate (FAR): percentage of images in $P_N$ that are falsely accepted
# Closed-set Benchmark Datasets

<table>
<thead>
<tr>
<th>Dataset</th>
<th>#Cameras</th>
<th>#Persons</th>
<th>#Images</th>
<th>#Views</th>
</tr>
</thead>
<tbody>
<tr>
<td>VIPeR</td>
<td>2</td>
<td>632</td>
<td>1,264</td>
<td>2</td>
</tr>
<tr>
<td>ETHZ</td>
<td>1</td>
<td>146</td>
<td>8,555</td>
<td>1</td>
</tr>
<tr>
<td>i-LIDS</td>
<td>5</td>
<td>119</td>
<td>476</td>
<td>2</td>
</tr>
<tr>
<td>QMUL GRID</td>
<td>8</td>
<td>250</td>
<td>1,275</td>
<td>2</td>
</tr>
<tr>
<td>PRID2011</td>
<td>2</td>
<td>200</td>
<td>1,134</td>
<td>2</td>
</tr>
<tr>
<td>CUHK01</td>
<td>2</td>
<td>971</td>
<td>3,884</td>
<td>2</td>
</tr>
<tr>
<td>CUHK02</td>
<td>5 pairs</td>
<td>1,816</td>
<td>7,264</td>
<td>2</td>
</tr>
<tr>
<td>CUHK03</td>
<td>6</td>
<td>1,360</td>
<td>13,164</td>
<td>2</td>
</tr>
<tr>
<td>CAMPUS-Human</td>
<td>3</td>
<td>74</td>
<td>1,889</td>
<td>3</td>
</tr>
<tr>
<td>Market-1501</td>
<td>6</td>
<td>1,501</td>
<td>32,668</td>
<td>-</td>
</tr>
<tr>
<td>MARS</td>
<td>6</td>
<td>1,261</td>
<td>1,191,003</td>
<td>-</td>
</tr>
<tr>
<td>DUKE</td>
<td>8</td>
<td>1,404</td>
<td>36,411</td>
<td>-</td>
</tr>
</tbody>
</table>
## Open-set Benchmark Datasets

<table>
<thead>
<tr>
<th>Dataset</th>
<th>#Cameras</th>
<th>#Persons</th>
<th>#Images</th>
<th>#Views</th>
</tr>
</thead>
<tbody>
<tr>
<td>Open-world</td>
<td>6</td>
<td>28</td>
<td>4,096</td>
<td>-</td>
</tr>
<tr>
<td>OPeRID</td>
<td>6</td>
<td>200</td>
<td>7,413</td>
<td>5</td>
</tr>
</tbody>
</table>
## Closed-set Benchmark Results

Benchmark on DukeMTMC-reID

<table>
<thead>
<tr>
<th>Methods</th>
<th>Rank@1</th>
<th>mAP</th>
</tr>
</thead>
<tbody>
<tr>
<td>BoW+kissme</td>
<td>25.13%</td>
<td>12.17%</td>
</tr>
<tr>
<td>LOMO+XQDA</td>
<td>30.75%</td>
<td>17.04%</td>
</tr>
<tr>
<td>PSE</td>
<td>79.8%</td>
<td>62.0%</td>
</tr>
<tr>
<td>ATWL(2-stream)</td>
<td>79.80%</td>
<td>63.40%</td>
</tr>
<tr>
<td>Mid-level Representation</td>
<td>80.43%</td>
<td>63.88%</td>
</tr>
<tr>
<td>HA-CNN</td>
<td>80.5%</td>
<td>63.8%</td>
</tr>
<tr>
<td>Deep-Person</td>
<td>80.90%</td>
<td>64.80%</td>
</tr>
<tr>
<td>MLFN</td>
<td>81.2%</td>
<td>62.8%</td>
</tr>
<tr>
<td>DuATM (Dense-121)</td>
<td>81.82%</td>
<td>64.58%</td>
</tr>
<tr>
<td>PCB</td>
<td>83.3%</td>
<td>69.2%</td>
</tr>
<tr>
<td>Part-aligned (Inception V1, OpenPose)</td>
<td>84.4%</td>
<td>69.3%</td>
</tr>
<tr>
<td>GP-reID</td>
<td>85.2%</td>
<td>72.8%</td>
</tr>
<tr>
<td>SPreID (Res-152)</td>
<td>85.95%</td>
<td>73.34%</td>
</tr>
</tbody>
</table>
Open-set Benchmark Results

- On OPeRiD

<table>
<thead>
<tr>
<th>Method</th>
<th>FAR=1% Rank=1</th>
<th>FAR=1% Rank=10</th>
<th>FAR=10% Rank=1</th>
<th>FAR=10% Rank=10</th>
</tr>
</thead>
<tbody>
<tr>
<td>IDENTITY</td>
<td>0.84</td>
<td>0.91</td>
<td>7.36</td>
<td>9.21</td>
</tr>
<tr>
<td>MAHAL [13]</td>
<td>1.89</td>
<td>1.99</td>
<td>10.50</td>
<td>11.97</td>
</tr>
<tr>
<td>LMNN [29]</td>
<td>0.41</td>
<td>0.41</td>
<td>3.97</td>
<td>4.58</td>
</tr>
<tr>
<td>ITML [6]</td>
<td>1.18</td>
<td>1.21</td>
<td>8.39</td>
<td>9.27</td>
</tr>
<tr>
<td>LADF [19]</td>
<td>1.53</td>
<td>1.74</td>
<td>9.11</td>
<td>10.82</td>
</tr>
<tr>
<td>RRDA</td>
<td><strong>3.99</strong></td>
<td><strong>4.35</strong></td>
<td><strong>14.51</strong></td>
<td><strong>16.72</strong></td>
</tr>
</tbody>
</table>

Very poor!

Future Directions

With the help of large datasets, deep learning methods have achieved much better performance, and are becoming more and more important for person re-identification.
Due to limited labeled data and large diversity in practical scenarios, semi-supervised learning or unsupervised learning will be potentially useful for practical applications in exploring large amount of unlabeled data.
Future Directions

Performance of cross-dataset evaluation is still poor. Unsupervised transfer learning and Re-ranking methods may be very useful in improving the performance.
Future Directions

For evaluation, open-set person re-identification and cross-dataset evaluation will be preferred in evaluating practical performance.
Thanks!

Shengcai Liao
Institute of Automation
Chinese Academy of Sciences

http://www.cbsr.ia.ac.cn/users/scliao/