Multi-modal Re-Identification: IOT + Vision for Residential Community Tracking

Dr. Kit Thambiratnam
General Manager, AI Center, The Seedland Group
Sept. 2020
Outline

- **Landscape** - Residential Communities in China
- **Problem** – Why tracking in communities?
- **Multi-modal tracking** - framework and results
  - Tracklets
  - Identity Stamping
  - Identity Conflation
- **Conclusions**
Residential Communities in China

Residential communities focus on the needs of households, providing secure housing integrated with recreation, groceries, retail and lifestyle services.
For 14 years, the Seedland Real Estate Group Co., Ltd. has been committed to the exploration and innovation of human science and technology in all aspects of life, connecting science and technology with humanities, and redefining human understanding of the relationship between themselves and living space.
Our goal is to pair smart **home**, smart **community**, signature smart premium **lifestyle services** and intelligent **logistics** to provide a **premier residential lifestyle experience**.

---

**Lifestyle Services**
- JoyHealth Service
- Remote Health
- Child Education
- Fresh Market (Vegetables/Meat)
- Intelligent Vending
- Cinema and Entertainment

---

**Intelligent Logistics**
- Self-driving Car
- Delight Delivery Robot

---

**Community**
- Elevator
- Camera
- Deliveries
- Access Control
- Parking

---

**Home**
- Hub
- Fridge
- Bingo
- Switches
- HVAC
- Curtains

---

**Human-Centric Intelligent Experiences**

---

**OneID | OneData | OneService**

Powered by Smart Life System
SLS Smart Community Solutions
Intelligent security, child safety, people flow analysis, intelligent delivery and contactless access control based on real-time IOT, cameras and AI.

SLS Intelligent Community Health
Intelligent health situational back-tracing for emergency contact tracing and multi-level risk alerts.

SLS Smart Business Services
Personalized and intelligent shopping through integrated community-wide ordering, standardized delivery, AI customer service assistance, intelligent business operations analysis.
Community Tracking
Why Community Tracking?

Community tracking delivers value to residents and property management while strictly preserving privacy and allowing user opt-out.

**Playground Tracing**
Only residents/approved visitors enter child areas.

**Delivery Tracing**
Delivery workers do not visit unnecessary areas.

**Prohibited Tracing**
Tracing rule breakers for property management.

**Contact Tracing**
Contact tracing for non-phone users (kids/elderly).

**Suspicious Tracing**
Retracing paths of suspicious behavior.

**Lost Item Tracing**
Retracing to find lost items, phones, toys, ...

© 2020 Seedland

No part of this presentation may be reproduced in whole or in part in any manner without the permission of Seedland.
Emergency Contact Back-Tracing
Combining Multiple Modalities

A user will interact with multiple devices and multiple modalities in their journey. Multi-modal allows us to see beyond cameras and generate semantic knowledge.

1. Family Structure
2. Travel Habits
3. Consumption
4. Health
5. Community Graph

Multi-Modal Tracking
- Detection (CenterNet '19)
- Vector Search (ANN)
- ReID (OSNet 'CVPR19)
- Face Attrib. (ArcFace CVPR19)
- IOT Conflation (Search)

Knowledge Graph
- Real-time knowledge fusion
- Real-time tagging (GBDT)
- Real-time graph (RT Graph)
- Relationship mining (Entropy)
# Challenges of Tracking In Residential Communities

<table>
<thead>
<tr>
<th>Face-Based Identity</th>
<th>Device-Based Identity</th>
<th>Privacy</th>
</tr>
</thead>
</table>
| • Bias - Face recognition is known to perform poorly on children and elderly. | • Mobile phones are NOT everywhere - children and elderly.  
• Bluetooth-tracking low penetration in China. | • Must support do-not-track.  
• Must support usage for intended purposes only. |

© 2020 Seedland

No part of this presentation may be reproduced in whole or in part in any manner without the permission of Seedland
User Consent and Tracking

Obtain **explicit consent** during smart gate registration

Consent agreement clearly lists **specific user-value features** – no blanket permission for tracking

**Non-ID associated** short-term tracking (eg. suspicious person back-tracing) are **implicitly agreed** to via **security notice**
Cross-Day Stable User ID Tracking

Global User ID Conflation

ID Stamping

Tracking

Key Challenges
- ID stable across days
- Real-time
- Cost-efficient
- Visual invariance (clothes, bags, hats, …)
- Population bias (children, elderly).

Leverage multiple modalities while real-time and cost efficient.
Tracking
Cross-Camera Tracking (REID)

Key idea: Recognize same person across cameras by searching a continuously updated database of previously-seen entities

OSNet\(^1\) multi-scale feature learning robust to scale, camera-distance

ID-Aware Tracking

Omni-Scale Embedding

Cross-Camera ReID Search

Tracklet ID - Assign to entire track (smooth per-frame noisy recognition)

<table>
<thead>
<tr>
<th>t</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>id</td>
<td>014</td>
<td>014</td>
<td>014</td>
<td>012</td>
<td>014</td>
<td>012</td>
</tr>
<tr>
<td>m</td>
<td>0.8</td>
<td>0.9</td>
<td>0.9</td>
<td>0.7</td>
<td>0.5</td>
<td>0.8</td>
</tr>
</tbody>
</table>

ID-aware tracking dramatically improves tracking MOTA and high FPS

CenterNet (detection) + Kalman (trajectory) + Embedding Distance (identity)

Seen Entities

Detected new entities Add new exemplars for known entities

1Zhou, Kaiyang and Yang, Yongxin and Cavallaro, Andrea and Xiang, Tao, “Omni-Scale Feature Learning for Person Re-Identification”, in The IEEE International Conference on Computer Vision (ICCV), 2019
Cross-Camera Tracking (REID)

Key idea: Recognize same person across cameras by searching a continuously updated database of previously-seen entities

ID-aware tracking dramatically improves tracking MOTA and high FPS

CenterNet (detection) + Kalman (trajectory) + Embedding Distance (identity)

OSNet\(^1\) multi-scale feature learning robust to scale, camera-distance

Tracklet ID - Assign to entire track (smooth per-frame noisy recognition)

<table>
<thead>
<tr>
<th>t</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>id</td>
<td>014</td>
<td>014</td>
<td>014</td>
<td>012</td>
<td>014</td>
<td>012</td>
</tr>
<tr>
<td>m</td>
<td>0.8</td>
<td>0.9</td>
<td>0.9</td>
<td>0.7</td>
<td>0.5</td>
<td>0.8</td>
</tr>
</tbody>
</table>

\(1\) Zhou, Kuiyang and Yang, Yongxin and Cavallaro, Andrea and Xiang, Tao, “Omni-Scale Feature Learning for Person Re-identification”, in The IEEE International Conference on Computer Vision (ICCV), 2019
Tracking Tasks and Public Benchmarks

**Single-Camera Tracking** – Detect and track a person within one camera

**Cross-Camera Tracking (REID)** – Identify the same user across different cameras

---

### Multiple Object Tracking Benchmark

<table>
<thead>
<tr>
<th>Method</th>
<th>Benchmark Set Top-1 Accuracy</th>
<th>Model Memory Size</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Market1501</td>
<td>DukeMTMC</td>
</tr>
<tr>
<td>Seedland</td>
<td>95.1</td>
<td>88.6</td>
</tr>
<tr>
<td>PCB</td>
<td>93.8</td>
<td>83.3</td>
</tr>
<tr>
<td>BFE-Net</td>
<td>94.0</td>
<td>88.9</td>
</tr>
<tr>
<td>DG-Net</td>
<td>94.8</td>
<td>83.6</td>
</tr>
</tbody>
</table>

© 2020 Seedland

No part of this presentation may be reproduced in whole or in part in any manner without the permission of Seedland
Tracking is critical for down-stream understanding

Learnings from ACM2020 Grand Challenge

• #1 position in the ACM 2020 Multimedia Grand Challenge for Large-scale Human-centric Video Analysis international competition.

• High ranked in down-stream intelligence (pose tracking, action recognition) primarily because of improved dense crowd tracking

<table>
<thead>
<tr>
<th>ID-aware tracking</th>
<th>Omni-scale embed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multi-Stage model ensemble</td>
<td>Perspective/occlude augment</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Team Name</th>
<th># Institution</th>
<th>MOTA</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st</td>
<td>Adaptive FairMOT</td>
<td>iSEE-SYSU &amp; ACCUVISION</td>
</tr>
<tr>
<td>2nd</td>
<td>JiaRen.AI</td>
<td>Seeland</td>
</tr>
<tr>
<td>3rd</td>
<td>Crowd-Tracker</td>
<td>Xidian University</td>
</tr>
<tr>
<td>Try private</td>
<td>Amazon</td>
<td></td>
</tr>
<tr>
<td>NewTracker</td>
<td>Tencent</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Team Name</th>
<th># Institution</th>
<th>w_AP@avg</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st</td>
<td>Seedland</td>
<td>57.5091</td>
</tr>
<tr>
<td>2nd</td>
<td>ccc</td>
<td>YITu &amp; National University of Singapore</td>
</tr>
<tr>
<td>3rd</td>
<td>DH_LIBA</td>
<td>Dahua Technology</td>
</tr>
<tr>
<td>JDAI</td>
<td>JD Tech &amp; UESTC</td>
<td>55.0193</td>
</tr>
<tr>
<td>WanDan</td>
<td>UESTC</td>
<td>54.5262</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Team Name</th>
<th># Institution</th>
<th>mAP@avg</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st</td>
<td>MSF</td>
<td>YITu &amp; National University of Singapore</td>
</tr>
<tr>
<td>2nd</td>
<td>VM</td>
<td>Seedland</td>
</tr>
<tr>
<td>3rd</td>
<td>CF</td>
<td>City University of Hong Kong</td>
</tr>
<tr>
<td>8A</td>
<td>Tencent</td>
<td></td>
</tr>
<tr>
<td>only_person_rgb</td>
<td>Sun Yat-sen University</td>
<td></td>
</tr>
</tbody>
</table>

© 2020 Seedland

No part of this presentation may be reproduced in whole or in part in any manner without the permission of Seedland.
Cross-Camera Tracking (ReID)

**Key idea:** Recognize same person across cameras by searching a continuously updated database of previously-seen entities.

- **Camera Group**
- **ID-Aware Tracking**
- **Omni-Scale Embedding**
- **OSNet**\(^1\) multi-scale feature learning robust to scale, camera-distance

**Cross-Camera ReID Search**

**Tracklet ID - Assign to entire track**
(smooth per-frame noisy recognition)

<table>
<thead>
<tr>
<th>t</th>
<th>id</th>
<th>m</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>014</td>
<td>0.8</td>
</tr>
<tr>
<td>3</td>
<td>014</td>
<td>0.9</td>
</tr>
<tr>
<td>4</td>
<td>014</td>
<td>0.9</td>
</tr>
<tr>
<td>5</td>
<td>012</td>
<td>0.7</td>
</tr>
<tr>
<td>6</td>
<td>014</td>
<td>0.5</td>
</tr>
<tr>
<td>7</td>
<td>012</td>
<td>0.8</td>
</tr>
</tbody>
</table>

**Seen Entities**

<table>
<thead>
<tr>
<th>id</th>
<th>camera</th>
<th>exemplars</th>
<th>m</th>
</tr>
</thead>
<tbody>
<tr>
<td>013</td>
<td>c41</td>
<td>[ ... ]</td>
<td>0.8</td>
</tr>
<tr>
<td>013</td>
<td>c44</td>
<td>[ ... ]</td>
<td>0.9</td>
</tr>
</tbody>
</table>

**Zhou, Kaiyang and Yang, Yongxin and Cavallaro, Andrea and Xiang, Tao, "Omni-Scale Feature Learning for Person Re-Identification", in The IEEE International Conference on Computer Vision (ICCV), 2019**

- **ID-aware tracking** dramatically improves tracking MOTA and high FPS
- **CenterNet (detection) + Kalman (trajectory) + Embedding Distance (identity)**
- **Key idea:** Recognize same person across cameras by searching a continuously updated database of previously-seen entities
ReID Search – Efficient Vector Search

KD-Tree Search – Efficient large database vector search
Partition tree at N level using (N mod k)th vector dimension

Cross-Camera ReID Search
Tracklet ID Assign

seen Entities

seen Entity Update

Detect new entities
Add new exemplars for known entities

Tracklet ID - Assign to entire track
(smooth per-frame noisy recognition)

<table>
<thead>
<tr>
<th>t</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>id</td>
<td>014</td>
<td>014</td>
<td>014</td>
<td>012</td>
<td>014</td>
<td>012</td>
</tr>
<tr>
<td>m</td>
<td>0.8</td>
<td>0.9</td>
<td>0.9</td>
<td>0.7</td>
<td>0.5</td>
<td>0.8</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>id</th>
<th>camera</th>
<th>exemplars</th>
<th>m</th>
</tr>
</thead>
<tbody>
<tr>
<td>013</td>
<td>c41</td>
<td>[ ... ]</td>
<td>0.8</td>
</tr>
<tr>
<td>013</td>
<td>c44</td>
<td>[ ... ]</td>
<td>0.9</td>
</tr>
</tbody>
</table>

from DataScienceCentral

© 2020 Seedland

No part of this presentation may be reproduced in whole or in part in any manner without the permission of Seedland
ReID Search Optimization

- Tracking requires **New IDs** to propagate in real-time across camera groups.
- **Multiple approaches** to balance **Recall** vs. **Compute Cost** vs. **Complexity**.
- Near-exchange provides best balance for community

<table>
<thead>
<tr>
<th>Approach</th>
<th>Max IDs</th>
<th>REID Search</th>
<th>Seen Entity Exchange</th>
<th>Latency</th>
<th>Recall</th>
<th>Complexity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Node-Inner</td>
<td>N</td>
<td>Linear Scan</td>
<td>None</td>
<td>Real-Time</td>
<td>Poor (mitigate by heavy nodes more cameras/node)</td>
<td>Trivial</td>
</tr>
<tr>
<td>Don’t exchange IDs</td>
<td></td>
<td>(MatMult)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Global Exchange</td>
<td>NxK</td>
<td>ANN Search</td>
<td>Periodic ANN Build</td>
<td>&gt;30s</td>
<td>Good but with latency</td>
<td>Complex</td>
</tr>
<tr>
<td>All nodes get IDs</td>
<td></td>
<td>(NxK large)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Near Exchange</td>
<td>NxV</td>
<td>Linear Scan</td>
<td>Real-time</td>
<td>Real-Time</td>
<td>Good</td>
<td>Near-Trivial</td>
</tr>
<tr>
<td>Only physically near</td>
<td></td>
<td>(MatMult)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>nodes get IDs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Global + Near</td>
<td>NxK</td>
<td>ANN + Linear Scan</td>
<td>Periodic ANN Build +</td>
<td>Real-Time</td>
<td>Best</td>
<td>More Complex</td>
</tr>
<tr>
<td>Best of both</td>
<td></td>
<td>(delta)</td>
<td>Real-time (delta)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[ N = \text{Expected Max \# People / Node} \quad K = \text{Number Nodes} \quad V = \text{Propagate to V-nearest nodes (complexity vs. recall trade-off)} \]
Community Tracking Challenges

Similar clothing

Workers all look the same

Old analogue cameras
Multi-Modal Identity Stamping
Face Identity Challenges

Easy

- Frontal face, good lighting.
- LFFW
  - Seedland: 99.8
  - Tencent: 99.8
  - Baidu: 99.8
  - Dahua: 99.8

Challenging

- Population bias e.g. Students
  - Seedland: 95.5@1e-6
  - Tencent: 99.8
  - Baidu: 99.8
  - Dahua: 99.8
  - SenseTime '18: 94.2@1e-6

Real Community Data

- Lighting, angle, expression, occlusion, blur very adverse! ~92% accuracy

© 2020 Seedland

No part of this presentation may be reproduced in whole or in part in any manner without the permission of Seedland
A typical user journey through the residential community has multiple touch points where user interacts with a device.

Multi-Modal identity stamping uses device interaction + visual features to convert the difficult identity task into an easier verification task.

1. Use ReID to build cross-camera tracklet
2. Assign ID at key device interaction points
3. Stamp ID on tracklet after disambiguating disagreements

Examples of device Interactions in a user journey
- Face access systems
- Bluetooth pairing
- Swipe card access systems
- Shop POS payment
POS data for Multi-Modal Association

<table>
<thead>
<tr>
<th>Camera Distance/Angle</th>
<th>Recall</th>
<th>Precision</th>
</tr>
</thead>
<tbody>
<tr>
<td>Face ID only</td>
<td>55%</td>
<td>63%</td>
</tr>
<tr>
<td>Ceiling - 3m/20°</td>
<td>52%</td>
<td>90%</td>
</tr>
<tr>
<td>POS camera – 0.5m/10°</td>
<td>30%</td>
<td>93%</td>
</tr>
</tbody>
</table>

Convert hard ID task to easy Verification task

- Synchronize POS+Camera events
- Verify instead of ID face
- Store Features as ReID Exemplar for future tracking
- Generate Association Record

Increases Precision by +30% abs. over face ID. Key challenges is disambiguation – who is paying, who is waiting.
Retail Shop Community Tracking

Smart Life System

实时热力图

顾客统计

顾客年龄

顾客性别

进店顾客购买统计

客单价统计

会员

非会员(CV)

非会员(订单)

运营统计

人数

实时客流量

© 2020 Seedland
No part of this presentation may be reproduced in whole or in part in any manner without the permission of Seedland.
POS data for Multi-Modal Association

Cross-Camera Tracking

Swipe Door Card

Door Access Data

Association Record

FaceFeature ➔ (userID, score)
REIDFeature ➔ (userID, score)

Multi-Modal Identity Stamping

Face feature

Body feature

Propagate User ID from Door Access to Visual Features

• Accumulate visual features that historically co-occur with card ID
• 1:N identification task, N is very small (<10)

CardID ➔ HouseID

Card is not unique to a person, people may give cards to family members, visitors, ....
Global User ID Conflation
Cross-Day Stable User ID Tracking

ID stamping User IDs are system-unique, not global.

Real-Time User ID Conflation **probabilistically associates** User IDs into a global User ID space that is stable and unique per user.
Real-Time User ID Conflation

Retrieval-based conflation does real-time assignment of a Multi-Modal Query event (text+visual) to a known set of stable Global User IDs. Multi-modal increases recall by 20% at same precision.

Query Event
which user assign query to

L1 Search
Text-search, High Recall

L2 Multi-Modal Reranking
Visual+Text-search, Top-1 Accuracy

L1 Candidates

L2 Re-reranked Candidates

Choose Top-1 UserID by Rank

Visual Rank

Match Rank

Query Features

Candidate Features

Visual features

Text features

Visual features

Text features

User ID
System Primary User IDs
Strong ID Features
phoneNumber
address
licensePlate

Visual Features
faceEmbed
bodyEmbed
visual attributes (age, sex)

Query
System Primary User IDs
Strong ID Features
phoneNumber
address
licensePlate

Visual Features
faceEmbed
bodyEmbed
visual attributes (age, sex)

L1 Search

L2 Multi-Modal Reranking

Choose Top-1 UserID by Rank

Visual Rank

Match Rank

Query Features

Candidate Features

Visual features

Text features

Visual features

Text features

User ID
System Primary User IDs
Strong ID Features
Visual Features

© 2020 Seedland

No part of this presentation may be reproduced in whole or in part in any manner without the permission of Seedland
Real-Time User Tracking and ID Conflation
Conclusions
Conclusions

Residential communities have many opportunities to transform community, lifestyle and retail with AIOT.

Multi-modal tracking is a vehicle for intelligent user experiences with significant benefits over vision-only.

Privacy and true value for residents must be a first-class citizen when playing in the tracking space.