Person Tracking

Localising people over time

Targeting indoor market research applications
- How many customers are there?
- What parts of the shop are they in?
- Are they going past high value items?
- Do we have enough staff?
Primary Challenges

Computational Efficiency
- Most approaches too slow for real-time
- Balance accuracy vs speed trade-off
- Hard constraint in embedded systems

Unsupervised Learning
- Person re-identification across cameras
- Lack of training data at execution time
- Often low-quality cameras, low detail
- Plug-and-Play approach instead of end-to-end CNN
- Modular approach better for distributed processing
- Reduced memory and bandwidth requirements
- Supports privacy-affirming framework: potentially, no human ever sees the raw footage
A Distributed Privacy-Affirming Architecture
Background Estimation: SuperBE

- Superpixel-based Background Estimation
- Isolate foreground as region of interest
- Minimise unnecessary processing later in pipeline
Person Detection: DPM

- Deformable Parts Model
- Isolates out head, torso, arms, legs
- Helps us deal with obscured parts of the body
Feature Extraction

- Extract colour and texture features for each part
- Use HSL for a cylindrical colour space
- Use LBP for computationally efficient texture descriptor
- Unsupervised sequential k-means model

1) Similarity and Classification

Use correlation distance to match histograms
Determine similarity of each part, ignoring obscured
Weight colour and texture features

\[ S = \alpha_1 \frac{\sum_p d_{col_p} v_p}{\sum_p v_p} + \alpha_2 \frac{\sum_p d_{lp_p} v_p}{\sum_p v_p} \]

Determine class with highest similarity
Feature Matching

- Unsupervised sequential k-means model

2) Model Update

We only maintain a class mean, not an entire cluster.

Modify the class mean with the new sample

\[ m_c = \beta x + (1 - \beta) m_c \]

Newer samples better indication of current person appearance than older samples.

Need some robustness against noise/false positives.
- Unsupervised sequential k-means model

3) Feature Weighting

Try to minimise inter-class similarities
Try to maximise inter-class variation
Help achieve better discriminability

Suppress the common background
Exaggerate the different foreground
Weight the values in the feature vectors
3) Feature Weighting

As each sample comes in, modify the weight vector
Apply weight vector during classification
Unsupervised learning, improves over time
- Assign the class ID to the detected person
- Detect position by taking midpoint between feet parts
- Form a track over time by connecting position points
- Create a test environment, similar to an office space
- Humans approximate zones, not exact co-ordinates
- Loosens requirements on precision, better comparison
Dataset

- Four cameras, partly overlapping, different angles
- Cheap webcams to emulate real-world video capture
- Seven action categories: walking, sitting, groups, etc.
Annotation Tool

- Uses homographies for real-world co-ordinates and zones
- Use optical flow to predict boxes, reduce annotation time
- Annotation tool released as open-source on Github
- New identity-based metrics for person tracking
- Furniture blocking cameras significantly drops accuracy
- Comparable to state-of-the-art re-identification: 60-80%
- Person detection is the bottleneck
- Each additional person requires approx. 50ms more
- Pipeline processes video with between 5-10fps
Summary

1. Computationally efficient person tracking
2. New online unsupervised learning approach to feature matching in real-time
3. Development of dataset and annotation tool
4. Development of pipelined system architecture to support future work
Future Work

1. Development of retail test scenario
2. Comparison with other re-identification and classification methods
3. Combine multiple camera views to localise position with high accuracy
4. Implementing distributed image processing architecture with smart cameras

Contact: andrew.chen@auckland.ac.nz