CMU-Informedia @ TRECVID 2011
Surveillance Event Detection

Longfei Zhang, Lu Jiang, Lei Bao, Shohei Takahashi, Yuanpeng Li,
Alexander Hauptmann
Carnegie Mellon University
SED11 Team

- Team members:

  Longfei  
  Lu  
  Lei  
  Shohei  
  Yuanpeng  
  Alex
Outline

- Framework
- MoSIFT based Action Recognition
  - MoSIFT feature
  - Spatial Bag of Word
  - Tackling highly imbalanced datasets
- Experiment Results
Framework

• Augmented Boosted Cascade
Framework

• Augmented Boosted Cascade
MoSIFT

- Given pairs of video frames, detect spatio-temporal interest points at multiple scales.
  - SIFT point detection with sufficient optical flow.
  - Describing SIFT points through SIFT descriptor and optical flow.
Spatial Bag of Words

- Each frame is divided into a set of non-overlapping rectangular tiles.
- The resulting BoW features are derived by concatenating the BoW features captured in each tile.
- Encode the spatial (tile) information in BoW.

Diagram:
- (Camera 3)
- Hot region detection
- K-means Clustering
  - K = 3000
- Feature quantization
- Spatial Bag of Words
- + + +
Hot Region Detection

• Person Detection: Person detection based on Histogram of Oriented Gradient (HOG) features.
• Background subtraction.

Over generated Person detection results for tracking and feature selection
Spatial Bag of Features

- Each frame is divided into a set of rectangular tiles or grids.
- The resulting Bow features are derived by concatenating the BoW features captured in each grid.
- Encode the adjusted spatial information in BoW.
Spatial Bag of Features

- Each frame is divided into a set of rectangular tiles or grids.
- The resulting BoW features are derived by concatenating the BoW features captured in each grid.
- Encode the adjusted spatial information in BoW.
Tackling the highly imbalanced data

- Augmented Cascade SVM.
- Bagging classification method except it adopts probabilistic sampling to select negative samples in a sequential manner.
Tackling the highly imbalanced data

- Augmented Cascade SVM.
- Bagging classification method except it adopts probabilistic sampling to select negative samples in a sequential manner.
Tackling the highly imbalanced data

- Augmented Cascade SVM.
- Bagging classification method except it adopts probabilistic sampling to select negative samples in a sequential manner.

<table>
<thead>
<tr>
<th>Training Dataset</th>
<th>Sub Dataset 1</th>
<th>Classifier 1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.8</td>
<td>0.7</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>0.3</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>0.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.1</td>
</tr>
</tbody>
</table>
Tackling the highly imbalanced data

• Augmented Cascade SVM.
• Bagging classification method except it adopts probabilistic sampling to select negative samples in a sequential manner.
Tackling the highly imbalanced data

- Augmented Cascade SVM.
- Bagging classification method except it adopts probabilistic sampling to select negative samples in a sequential manner. N = 10 layers.
Tackling highly imbalanced data

Bagging Ensemble of Random Forests

• Random Forest is a forest of decision trees.

• Two parameters:
  – n is the number of trees in the forest.
  – m the number of features in each decision tree.

• Build each decision tree by randomly selecting m features and use C4.5.

• Each tree is grown without pruning.
Tackling highly imbalanced data
Bagging Random Forest: Ensemble of Random Forests

- Random Forest is a forest of decision trees.
- Two parameters:
  - $n$ is the number of trees in the forest.
  - $m$ is the number of features in each decision tree.
- Builds each decision tree by randomly selecting $m$ features.
- Each tree is grown without pruning.
## Cascade SVM vs. Bagging Random Forest

<table>
<thead>
<tr>
<th></th>
<th>Cascade SVM (chi² kernel)</th>
<th>Bagging Random Forest</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Effectiveness</strong></td>
<td>Most Effective</td>
<td>Usually 3-8% less in Average Precision</td>
</tr>
<tr>
<td><strong>Efficiency</strong></td>
<td>Time consuming</td>
<td>Usually tens to hundreds of times faster</td>
</tr>
<tr>
<td><strong>Sensitive to Parameter settings</strong></td>
<td>Sensitive</td>
<td>Relatively insensitive</td>
</tr>
</tbody>
</table>

### Performance Comparison

![Bar Chart](image.png)

- **Training Time**:
  - Bagging Random Forest: 10 times faster
  - Cascade SVM: 94.5 times faster

- **Testing Time**:
  - Bagging Random Forest: 10 times faster
  - Cascade SVM: 94.5 times faster
Results

- 8 Submissions:
  - The first 6 runs use cascade SVM with different sliding window sizes and parameter sets.
  - Last 2 runs use bagging random forest method.
Results

- Results for **Primary** run:

<table>
<thead>
<tr>
<th></th>
<th>Inputs</th>
<th>Actual DCR</th>
<th>Minimum DCR</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>#Targ</td>
<td>#NTarg</td>
<td>#Sys</td>
</tr>
<tr>
<td>CellToEar</td>
<td>194</td>
<td>127</td>
<td>128</td>
</tr>
<tr>
<td>Embrace</td>
<td>175</td>
<td>657</td>
<td>715</td>
</tr>
<tr>
<td>ObjectPut</td>
<td>621</td>
<td>57</td>
<td>58</td>
</tr>
<tr>
<td>PeopleMeet</td>
<td>449</td>
<td>336</td>
<td>381</td>
</tr>
<tr>
<td>PeopleSplitUp</td>
<td>187</td>
<td>115</td>
<td>118</td>
</tr>
<tr>
<td>PersonRuns</td>
<td>107</td>
<td>413</td>
<td>439</td>
</tr>
<tr>
<td>Pointing</td>
<td>1063</td>
<td>1960</td>
<td>2092</td>
</tr>
</tbody>
</table>
Results

Compared with our primary run with those of other teams.
We have the best Min DCR in 3 out of 6 events.
Results

Compared with our primary run with those of other teams. We have the best Actual DCR in 3 out of 7 events.
Results

Compared with our last year’s result, we get improvement in terms of MIN DCR in 5 events “Embrace”, “People Meet”, “People Split up”, “Person Runs” and “Pointing”.

- Best event results over all CMU runs

<table>
<thead>
<tr>
<th>Min DCR</th>
<th>Cell To Ear</th>
<th>Embrace</th>
<th>Object Put</th>
<th>People Meet</th>
<th>People Split Up</th>
<th>Person Runs</th>
<th>Pointing</th>
</tr>
</thead>
<tbody>
<tr>
<td>2010 CMU</td>
<td>1.0003</td>
<td>0.9838</td>
<td>1.0003</td>
<td>0.9793</td>
<td>0.9889</td>
<td>0.9477</td>
<td>1.0003</td>
</tr>
<tr>
<td>2010 Overall Best Event</td>
<td>1</td>
<td>0.9663</td>
<td>0.9971</td>
<td>0.9787</td>
<td>0.9889</td>
<td>0.6818</td>
<td>0.996</td>
</tr>
<tr>
<td>2011 CMU</td>
<td>1.0003</td>
<td>0.8658</td>
<td>1.0003</td>
<td>0.9684</td>
<td>0.7838</td>
<td>0.837</td>
<td>0.9996</td>
</tr>
</tbody>
</table>
Results

Compared with the best event results in TRECVID 2010, for event “Embrace”, “PeopleMeet” and “People Split Up” ours are the best system.

<table>
<thead>
<tr>
<th>Min DCR</th>
<th>Cell To Ear</th>
<th>Embrace</th>
<th>Object Put</th>
<th>People Meet</th>
<th>People Split Up</th>
<th>Person Runs</th>
<th>Pointing</th>
</tr>
</thead>
<tbody>
<tr>
<td>2010 CMU</td>
<td>1.0003</td>
<td>0.9838</td>
<td>1.0003</td>
<td>0.9793</td>
<td>0.9889</td>
<td>0.9477</td>
<td>1.0003</td>
</tr>
<tr>
<td>2010 Overall Best Event</td>
<td>1</td>
<td>0.9663</td>
<td>0.9971</td>
<td>0.9787</td>
<td>0.9889</td>
<td>0.6818</td>
<td>0.996</td>
</tr>
<tr>
<td>2011 CMU</td>
<td>1.0003</td>
<td>0.8658</td>
<td>1.0003</td>
<td><strong>0.9684</strong></td>
<td><strong>0.7838</strong></td>
<td>0.837</td>
<td><strong>0.9996</strong></td>
</tr>
</tbody>
</table>
Cascade SVM vs. Random Forest

- Comparison between Run 1 (Cascade SVM) and Run 7 (Random Forest) in terms of Min DCR.
Threshold Search

• Searching for Min DCR using cross validation.
• Actual DCR provides reasonable estimates of Min DCR on all runs.
Impact of sliding window size

- Results for all events with sliding window size 25 frames (Run 3).
Impact of sliding window size

- Results for all events with sliding window size 60 (Run 5).
Event-specific sliding window size

- For PersonRuns, CellToEar, Embrace and Pointing a good sliding window is small.
- For Embrace, ObjectPut and PeopleMeet a good sliding window size is larger.
Conclusions

- Observations:
  - MoSIFT feature captures salient motions in videos.
  - Spatial Bag of Words can boost the performance over last year’s result.
  - Event-specific sliding window size impacts the final result.
  - Both cascade SVM and bagging random forest can handle highly imbalanced data sets. Random forest is much faster.
THANK YOU.
Q&A?