Exemplar-SVM for Action Recognition

Christina Peterson, Rui Hou
(clp8199@knights.ucf.edu, houray@gmail.com)
University of Central Florida

1. Problem
- Objective: Action Recognition in Videos Using a Series of Exemplar-SVMs
- Motivation: Only one labeled video as a positive sample is required to train an Exemplar-SVM, reducing the amount of required labeled data.
- Allows for training with very little human supervision.

2. UCF Sports Actions Dataset
- Set of ten actions collected from various sports which are typically featured on broadcast television channels.
- Each exemplar is trained against the other actions in the training dataset.

3. Overall Approach
(1) Bag of Words
- Collect Low Level Features within Bounding Box
- STIP (Space-Time Interest Points)
- Histogram of Oriented Gradient (HOG)
- Histogram of Optical Flow (HOF)
- Color
- Texture
- Cluster by K-means
- 128 for Color, 256 for Texture, 1000 for STIP
- Construct Feature Vector
- Create histogram of cluster centers per feature for each cell in bounding box
- Normalize Histograms
- Final Feature Vector is (128 + 256 + 1000) x 10 = 13,840 dimensions

(2) Train Exemplar-SVMs
- Select a minimal number of videos from each action class which serve as a good representation for the action.
- Train a series of Linear-SVMs where the selected exemplar video is the positive sample and videos from other actions are the negative samples.
- Negative samples also include background volumes.

(3) Calibration
The calibration step is performed in order to make the raw scores of the Exemplar-SVMs compatible.
- Platt’s Method
- Fit a probability distribution to a validation set
  \[ P(x) = \frac{1}{1 + e^{-\alpha(x-\beta)}} \]

(4) Calibrated Exemplar-SVM Scores as Mid-Level Features
All Exemplar-SVMs are tested on the validation set, and the calibrated scores of all exemplars are concatenated together to form a mid-level feature vector for the validation video.

(5) Train Multi-Class SVM based on Calibrated Exemplar-SVM Scores
A Multi-Class SVM is trained on the mid-level features vectors of the validation set.

4. Experimental Evaluation
Testing was performed on the UCF Sports Actions Dataset.
- Exemplar-SVM set was held consistent.
- The remaining videos from each action class were used to train and test the Multi-Class SVM.
- The Multi-Class SVM was trained using a two-fold cross validation method.
- Videos split into two groups.
- First group used for training, second group used for testing.
- Second group used for training, first group used for testing.
- Results between test cases were averaged.

5. Results
- Recognition Accuracies on UCF Sports Actions Dataset

<table>
<thead>
<tr>
<th>Method</th>
<th>Diving</th>
<th>Golf</th>
<th>Kick</th>
<th>Lift</th>
<th>Run</th>
<th>Horse-Ride</th>
<th>Ski</th>
<th>Skate</th>
<th>Hang-Bench</th>
<th>Swing-Bench</th>
<th>Walk</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exemplar-SVM (Ours)</td>
<td>56.6</td>
<td>87.5</td>
<td>83</td>
<td>100</td>
<td>97</td>
<td>92</td>
<td>97</td>
<td>96</td>
<td>100</td>
<td>95</td>
<td>92</td>
</tr>
<tr>
<td>Wuer et al.</td>
<td>65.8</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>99</td>
<td>100</td>
<td>99</td>
</tr>
<tr>
<td>Action Bank</td>
<td>94.4</td>
<td>100</td>
<td>91.7</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>97</td>
<td>67</td>
<td>16.7</td>
</tr>
<tr>
<td>Wu et al.</td>
<td>91.3</td>
<td>100</td>
<td>88</td>
<td>100</td>
<td>100</td>
<td>67</td>
<td>93</td>
<td>84</td>
<td>91</td>
<td>91</td>
<td>88</td>
</tr>
<tr>
<td>Le et al.</td>
<td>95.0</td>
<td>100</td>
<td>100</td>
<td>83</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>94</td>
<td>90</td>
<td>92</td>
<td>94</td>
</tr>
</tbody>
</table>

Confusion Matrix for Exemplar-SVM