New Challenges in Semantic Concept Detection

National Taiwan University

Preliminaries for Semantic Concept Detection

• What are preliminaries for building a semantic concept detection system?
  – A lexicon of well-defined concepts
  – Training resources
    • Video data
    • Annotations
    • Features
  – Tools
    • Tagging or labeling tools, (e.g., CMU and IBM tools)
    • Feature extractors
    • Machine learning tools, (e.g., LIBSVM)
    • Semantic concept detection tailored tools
Semantic Concepts

TRECVID-2005
10 Concepts

TRECVID-2006
LSCOM-Lite (39)

MediaMill 101

Columbia374

LSCOM 449

Video Data Sets

Dataset

TV 07
TV 06
TV 05
TV 04
TV 03

Video Length (hrs)

0 50 100 150 200

100 hours
Sound and
Vision Video...

~190 hours
devel. set

~190 hours
test set

~330 hours
Multi-Lang.
Broadcast
News Video

~190 hours
News Video
### Features, Detectors, Scores

<table>
<thead>
<tr>
<th></th>
<th>Features</th>
<th>Detectors</th>
<th>Scores</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>MediaMill</strong></td>
<td>• Visual feature</td>
<td>5 sets of 101 classifiers</td>
<td>Scores of TV 05/06 dataset</td>
</tr>
<tr>
<td><strong>Baseline</strong></td>
<td>• Text feature</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Columbia374</strong></td>
<td>• EDH</td>
<td>Columbia 374 detectors</td>
<td>Columbia374 scores of TV 06/07 dataset</td>
</tr>
<tr>
<td></td>
<td>• GBR</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• GCM</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>VIREO-374</strong></td>
<td>• Color moment</td>
<td>VIREO-374 detectors</td>
<td>Scores of TV 07 dataset</td>
</tr>
<tr>
<td></td>
<td>• Wavelet texture</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Keypoint feature</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Available Resources

<table>
<thead>
<tr>
<th>Training Resources</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Well-defined concepts</td>
<td>✔️</td>
</tr>
<tr>
<td>Video data</td>
<td>✔️</td>
</tr>
<tr>
<td>Annotations</td>
<td>✔️</td>
</tr>
<tr>
<td>Features</td>
<td>✔️</td>
</tr>
<tr>
<td>Tools</td>
<td></td>
</tr>
<tr>
<td>Tagging Tools</td>
<td>✔️</td>
</tr>
<tr>
<td>Feature Extractors</td>
<td>✗</td>
</tr>
<tr>
<td>Machine Learning Tools</td>
<td>✔️</td>
</tr>
<tr>
<td>Tailored Tools</td>
<td>✗</td>
</tr>
</tbody>
</table>

• Concept definition is sufficient
• Training resources are plentiful
• No feature extractors and tailored tools available

The New Challenges

• **Challenge 1: Easy and Efficient Tools**
  – $L$ datasets, $M$ concepts, $N$ features, imply $L*M*N$ classifiers
  – Each classifier has to consider many parameters
  – **Time seems very limited** to validate each parameter and to train all classifiers

• **Challenge 2: Resource Exploitation or Reuse**
  – Resources are precious
  – Existent resources are **potentially useful** for new dataset
  – Plentiful resources have **not been fully utilized**
Facing the New Challenges

• **Challenge 1:**
  – Extended LIBSVM to improve training efficiency
  – Developed an efficient and easy-to-use toolkit tailored for semantic concept detection

• **Challenge 2:**
  – Reused classifiers of past data to improve accuracy by late aggregation
  – Exploited contextual relationship and temporal dependency from annotations to boost accuracy
A Tailored Toolkit

• We extended LIBSVM in three aspects for semantic concept detection:
  – Using dense representations
  – Exploiting parallelism of independent concepts, features, and SVM model parameters
  – Narrowing down parameter search to a safe range

• Overall, training time of our baseline was approximately reduced from 14 days to about 3 days

Facing the New Challenges

• Challenge 1
  – Extended LIBSVM to improve training efficiency
  – Developed an efficient and easy-to-use toolkit tailored for semantic concept detection

• Challenge 2: Resource Exploitation or Reuse
  – Reused classifiers of past data to improve accuracy by late aggregation
  – Exploited contextual relationship and temporal dependency from annotations to boost accuracy
Reuse Past Data

• Early aggregation
  – Must re-train classifiers
  – Cause considerable training time

• Late aggregation
  – Simple and direct
  – May be biased

Late Aggregation

• We adopt late aggregation to reuse existent classifiers by two strategies:
  – Equally Average Aggregation
    • Simply average the scores of past and newly trained classifiers
  – Concept-dependent Weighted Aggregation
    • Use concept-dependent weights to aggregate classifiers
### Aggregation Benefits

- **Overall Improvement Ratio** –
  - Average Aggregation: 22%
  - Weighted Aggregation: 30%

### Facing the New Challenges

**Challenge 1**
- Extended LIBSVM to improve training efficiency
- Developed an efficient and easy-to-use toolkit tailored for semantic concept detection

**Challenge 2: Resource Exploitation or Reuse**
- Reused classifiers of past data to improve accuracy by late aggregation
- Exploited contextual relationship and temporal dependency from annotations to boost accuracy
Observation in Annotations

A sequence of video shots

A lexicon of concepts

Temporal Dependency

Contextual relationship

Post-processing Framework

Feature Extraction → Concept Detection

Video Segmentation → Annotation

Unsupervised Contextual Fusion → Concept Reranking

Temporal Dependency Mining → Temporal Filter Design → Temporal Filtering

Shot Ranking

Combination

Detecting phase: Mining phase: Processing phase:
Temporal Filtering

- Feature Extraction
- Concept Detection
- Video Segmentation
- Annotation
- Video Sequence

Temporal Filtering

- Unsupervised Contextual Fusion
- Concept Reranking
- Shot Ranking
- Combination

Detecting phase: —— Mining phase: —— Processing phase: ——

Temporal Dependency

- Different concepts have different levels of dependency at different temporal distance
  - E.g., sports, weather, maps, explosion

Chi-square test

- sports
- weather
- maps
- explosion

Temporal Distance
Temporal Filter

\[ X_{t-k} \cdots X_{t-2} \rightarrow X_{t-1} \rightarrow X_t \rightarrow X_{t+1} \rightarrow X_{t+2} \rightarrow \cdots \]

\[ \text{SVM Classifier} \]

\[ \hat{P}(t = 1) = \sum_{k=0}^{D} w_k (P(t = 1 | x_{-1})) \]

Filtering Prediction

- A sequence of shots for predicting *sports*

- **Classifier prediction results**

- **After temporal filtering**

  - Rank rose
  - Misclassification picked up
Concept Reranking

Detecting phase: ———— Mining phase: ———— Processing phase: ————

Initial ranking list produced by a baseline method
Target concept: ‘boat’ (search or detection)
Step 1. Randomly split to training and test sets

Initial list

Step 2. Learn to maintain the ranking orders

Initial list

Shot pairs
1. Related concepts
2. Importance of each concept

Related concepts:
ocean
waterscape
Step 3. Context fusion on the test data

Step 4. Merge
Combination

Detecting phase:  Mining phase:  Processing phase:

Post-processing Benefits

Overall improvement ratio:
Parallel Combination: +10%

+45%
+37%
+29%
+17%
Conclusion

• We reduce the training time of detectors by using a tailored toolkit for semantic concept detection

• The proposed aggregation methods reuse the classifiers of past data and can boost the detection accuracy

• Our post-processing approaches exploit existent resource and can further improve detection results

Thank You For Your Attention