TNO 2013 approach to TRECVID MED

Klamer Schutte, Henri Bouma, George Azzopardi, Martijn Spitters, Joost de Wit, Corné Versloot, Remco van der Zon, Pieter Eendebak, Jan Baan, Johan-Martijn ten Hove, Adam van Eekeren, Frank ter Haar, Richard den Hollander, Jasper van Huis, Maaike de Boer, Gert van Antwerpen, Jeroen Broekhuijzen, Laura Daniele, Paul Brandt, Wessel Kraaij
GOOSE and TRECVID MED

- TNO MED submission part of the GOOSE project.

- We will discuss the TRECVID MED task as seen from the wider GOOSE perspective, argue how MED can model a simplified GOOSE system, and our GOOSE system used for TRECVID submission.

- Since this is our first year participation in MED, our objective as to build a baseline system.

- We designed our GOOSE-MED system taking advantage of successful proven strategies of MED 2012 participants.
GOOSE Challenge

- Day-to-day life dominated by Internet everywhere and instant knowledge of friends activity using social media
- Current Military Operations dominated by last century technology

- Many sensors Internet connected
  - A minority dedicated to military operations
  - Too much data to check

- User wants answers to his query, not lots of sensor data

- Web 1.0 made by Internet search engines

- Internet of Things needs new paradigm similar to keyword search for web pages

- Allow ISR chains to use all sensor data
  - And allow to exploit this data down to platoon level
GOOSE Goal

- The GOOSE (GOOgle for SEnsors) concept has the ambition to provide the capability to search semantically for any relevant information within “all” (including imaging) sensor streams, in near real time, in the entire internet of sensors. Similar to the capability provided by presently available search engines which enable the retrieval of information on “all” pages on the internet.
GOOSE Big Technology Issues

› **Scalability**
  › number of sensors;
  › number of users;
  › diversity of queries;
  › diversity of application domains

› **Semantic gap**
  › To translate user queries to sensor processing;
  › To translate processing results to answers for users

› **Also consider**
  › Security
  › Privacy
  › Payments
GOOSE basic architecture
Semantic Gap

Operational information needs

How can a user formulate a query effectively?

Man machine interaction

What domain knowledge is needed to interpret this question?
How to map specific information need to the generic processing?

Processing modules

What generic features can filter sensor data based on the information need?
How can we make specific verification with low bandwidth?

Sensor data

What sensors are needed for every question?
GOOSE and TRECVID MED

 › Basic design elements within GOOSE to close the semantic gap:
   1. using a semantic analysis of the user query
   2. use external crowdsourced knowledge sources, including semantic web, Imagenet, Google Images, Flickr, Youtube etcetera, to obtain specific understanding of domains not specifically considered at design (& learning) time of the system
   3. rely on user interaction to disambiguate concepts and indicate appropriateness of external crowdsources indicators.

 › Note that 2013 MED guidelines do not allow (2) and (3) design elements to close the semantic gap. We expect that truly open domain systems will need to use external data sources, and that in the short and medium term user interaction will be needed to disambiguate complex user queries and/or domain specifics.
Inspired by MED 2012
CMU SESAME ECNU
BBNVISER SRIAURORA
MediaMill AXES Tokyo GENIE IBM

All video data without label

Low level features

Feature Rep.

SVM apply event

SVM apply event

Late fusion event

Low level features

Feature Rep.

SVM apply concept

SVM apply concept

SVM train concept

SVM train concept

Event classifier

Semantic analysis

Google / YouTube With concept label

Google / YouTube With event label

AdHoc Labeled Event Video (10/100)

AdHoc Labeled Event Description (0/10/100)

Imagenet / Google Images / Youtube With concept label

Blue: TNO Additions
Applying TRECVID Pipeline to general GOOSE concept

Green area: First stage
- Sensor data
  - Low level features
  - Feature Rep.
  - SVM apply concept
  - SVM apply event
  - Late fusion event

Pink area: System config
- TNO approach to TRECVID MED
  - SVM apply concept
- All video data without label
  - AdHoc Labeled Event Video (10/100)

User Query
- Semantic analysis
- Imagenet / Google Images / Youtube With concept label
- Google / YouTube With event label

Not Available
- Low level features
  - Feature Rep.
  - SVM apply concept
  - SVM train event
  - SVM train concept
  - Event classifier
  - SVM apply event
  - SVM train event

Blue area: Sensor Independent
- Orange area: Final Stage
  - Application of TRECVID Pipeline to general GOOSE concept
TNO approach to TRECVID MED

SVM apply concept

All video data without label

AdHoc

Labeled

Event Video (10/100)

AdHoc

Labeled

Event Description (0/10/100)

Semantic analysis

Imagenet / Google Images / Youtube

With concept label

Low level features

Feature Rep.

SVM apply event

Late fusion event

SVM apply concept

SIFT

Sparse BoW (PCA)

SVM apply concept

SIFT

Sparse BoW (PCA)

SVM train event

Not Allowed

Not Available

Extremely limited setup to show GOOSE system concept!

Current MED limiting GOOSE: no online download!
Semantic Analysis flow

**Semantic Reasoning Component**

- **Lexical analysis**
- **Syntactical analysis**
- **Semantic interpretation**
- **Semantic analysis**
- **Query building**

**Event Kit**
- Evidential description

**OWL Ontology**
- [Textual concepts]
- [owl concepts]
- [WN synsets]

**WordNet Taxonomy**
- [Textual concepts derived from OWL concepts]

**High Level Concepts**
- [Textual concepts derived from WN synsets]
## Semantic analysis – example

**Win a race without a vehicle**

<table>
<thead>
<tr>
<th>Event Name</th>
<th>Winning a race without a vehicle</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Evidential Description</strong></td>
<td></td>
</tr>
<tr>
<td>scene</td>
<td>outdoors (park, field, track, road, or stadium) or indoors (indoor track, pool, or large gymnasium)</td>
</tr>
<tr>
<td>objects/people</td>
<td>runner, number worn on runner’s back/front/arm, potato sack, marker for finish line (tape stretched across road, potato sacks lying on ground), running shoes, baton, spectators, boundary markers/signs, signs supporting/encouraging a particular runner, water bottles, first aid tent</td>
</tr>
<tr>
<td>activities</td>
<td>running, swimming, hopping, climbing, jumping, breaking through tape, passing a baton, spectators running a short distance with the runner, passing out water bottles to the runners</td>
</tr>
<tr>
<td>audio</td>
<td>onlookers cheering, verbal or other indication of starting the race (yelling ”Go!”, gun shooting), narration of the race (speaking through a microphone)</td>
</tr>
</tbody>
</table>
Semantic analysis: AND of OR

Win a race without a vehicle

AND (racing(1)
    OR (NOT (vehicle (1), truck (1), tractor (1), car (1),
        bus (1), ambulance (1), policecar (1), taxi (1), boat (1),
        cruiseship (1), ship (1), sailingboat (1), rowingboat (1),
        motorboat (1), train (1), bicycle/bike (1), motorcycle (1),
        airplane (1), helicopter (1)))
    park (1)
    field (1)
    track (1)
    road (1)
    stadium (1)
    swimmingpool (1)
    runner (1)
    potato (1)
    finishline (1)
    tape (1)
    shoes (1)
    spectator (1)
    OR (water (1), food (0.69))
    bottle (1)
    sign(1)
    OR (tent (1), circumstent (1))
    run (1)
    swim (1)
    cheering (1)
    yelling(1)
    go (1)
    gun (1)
    shooting (1)
    person (0.3)
    microphone (1) )
Semantic Event Classifier

Applied to
- SIFT (418 concepts)
- LBP (442 concepts)
- MFCC (86 concepts)

Downloaded from Google Images and Youtube without human check

SVM scores normalized over training set

Weighted by
- semantic distance
- detectability value: average score of concept in training set where identified by semantic analysis
BoW setup

<table>
<thead>
<tr>
<th>Feature</th>
<th>Vocabulary size</th>
<th>Spatial tiling</th>
<th>Histogram size</th>
</tr>
</thead>
<tbody>
<tr>
<td>SIFT</td>
<td>300</td>
<td>Spatial pyramid: 1x1 + 2x2 + 4x4</td>
<td>6300</td>
</tr>
<tr>
<td>Opp. SIFT</td>
<td>300</td>
<td>Spatial pyramid: 1x1 + 2x2</td>
<td>1500</td>
</tr>
<tr>
<td>LBP</td>
<td>300</td>
<td>2x2</td>
<td>1200</td>
</tr>
<tr>
<td>STIP</td>
<td>300</td>
<td>3x3</td>
<td>2700</td>
</tr>
<tr>
<td>MFCC</td>
<td>300</td>
<td>N/A</td>
<td>300</td>
</tr>
</tbody>
</table>

Unexpected results:

- Vocabulary size of 300 outperformed 100, 1000 and 3000
- VLAD (in combination with PCA) didn’t improve performance
### Performance different features on training data

**Note:** MED Evaluation provided 2x3 numbers only!

<table>
<thead>
<tr>
<th>Method</th>
<th>Pre specified (PS)</th>
<th>Ad hoc (AH)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>pMiss @ TER (%)</td>
<td>pFA @ TER (%)</td>
</tr>
<tr>
<td>Semantic: MFCC</td>
<td>90.2</td>
<td>7.3</td>
</tr>
<tr>
<td>Semantic: LBP</td>
<td>85.6</td>
<td>6.5</td>
</tr>
<tr>
<td>Semantic: SIFT</td>
<td>79.9</td>
<td>6.2</td>
</tr>
<tr>
<td>Feature: STIP</td>
<td>79.3</td>
<td>6.4</td>
</tr>
<tr>
<td><strong>Fusion: 3 semantic (i)</strong></td>
<td>78.0</td>
<td>6.1</td>
</tr>
<tr>
<td>Feature: MFCC</td>
<td>74.0</td>
<td>6.0</td>
</tr>
<tr>
<td>Feature: LBP</td>
<td>65.7</td>
<td>5.2</td>
</tr>
<tr>
<td>Feature: Opponent-SIFT</td>
<td>59.1</td>
<td>4.7</td>
</tr>
<tr>
<td>Feature: SIFT</td>
<td>57.6</td>
<td>4.5</td>
</tr>
<tr>
<td><strong>Fusion: 5 features</strong></td>
<td>48.5</td>
<td>3.7</td>
</tr>
<tr>
<td><strong>Fusion: 5 features + 3 semantic (i)</strong></td>
<td>47.8</td>
<td>4.0</td>
</tr>
<tr>
<td>Feature: CC D-STIP 1FPS (not used)</td>
<td>68.4</td>
<td>±5.5</td>
</tr>
<tr>
<td>Concept: SVM SIFT (not used)</td>
<td>63.7</td>
<td>5.1</td>
</tr>
<tr>
<td>Fusion: 5 features trained using on-the-fly downloaded video (not allowed)</td>
<td>---</td>
<td>---</td>
</tr>
</tbody>
</table>
Performance features versus events

![Heatmap](image_url)
Discussion

› Entry barrier proved hard
  › Notebook papers of 2012 not sufficient for “fine” details

› Likely improvement areas
  › Temporal sampling
  › Dense features
  › Deep Learning
  › VLAD / Fisher vectors / …
  › Unbalanced data set & SVM
  › Concept detectors
  › Semantic representation
2014 MED TNO submission?

- Little to be gained
  - No major funding available – incremental change expected
  - No multiple run submission & evaluation
    - Allowing evaluation of different innovations
    - Allowing learning of different innovations tested by other team
  - Possible solution: shared obligatory submission on test set!

- Efforts not well aligned with GOOSE goals (& funding)
  - GOOSE semantic gap addresses user search goal <-> data
  - GOOSE scalability relies on external data sources
  - GOOSE scalability includes different users & domains
  - GOOSE verification stage not allowed in MED tasks