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 Broekhuijsen, 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.*

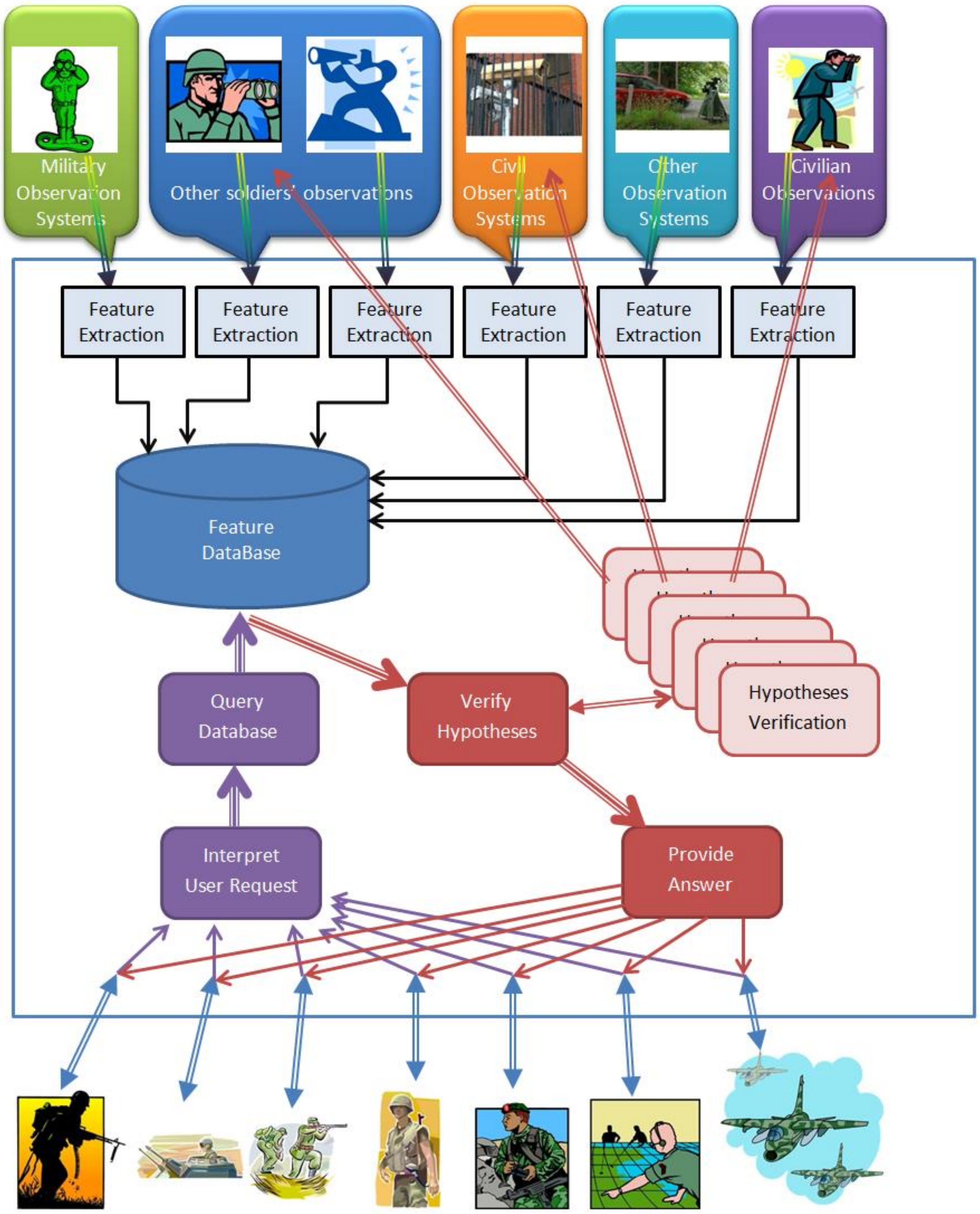
GOOSE Big Technology Issues

- ***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

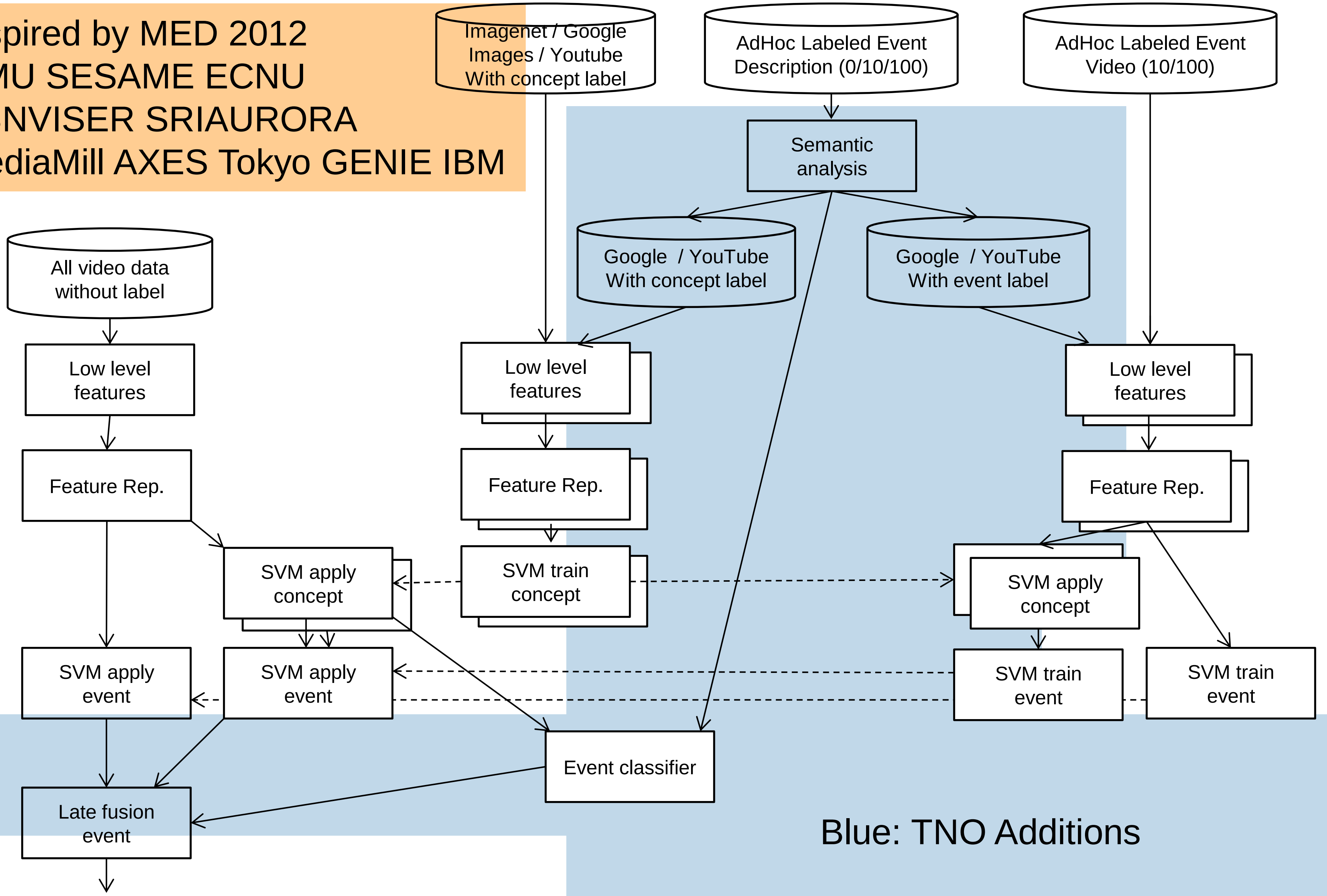
What sensors are needed for every question?

GOOSE and TRECVID MED

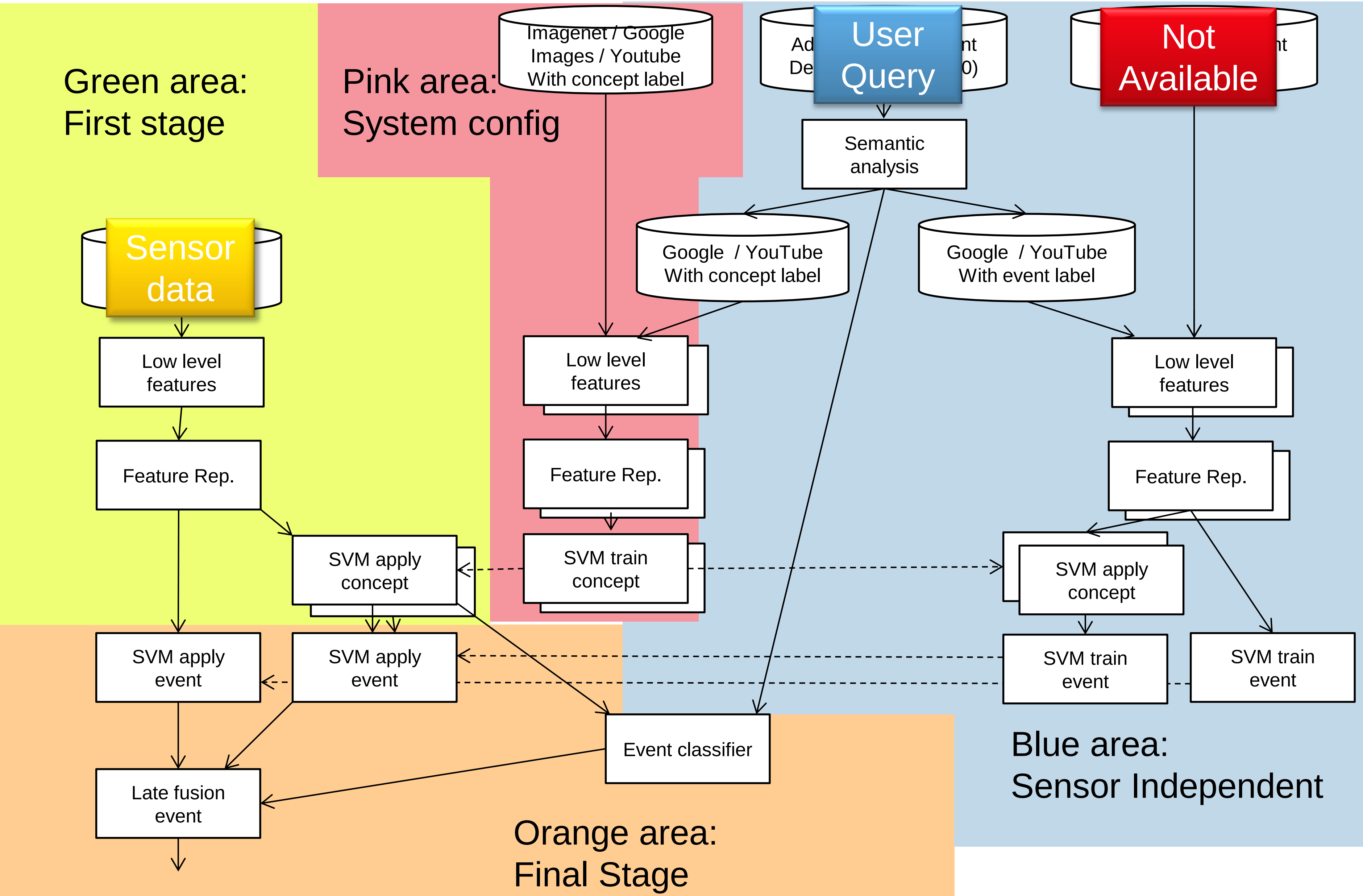
- Basic design elements within GOOSE to close the semantic gap:
    1. using a semantic analysis of the user query
    2. use external crowdsourced knowledge sources, including semantic web, Imagenet, Google Images, Flickr, Youtube etcetera, to obtain specific understanding of domains not specifically 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.

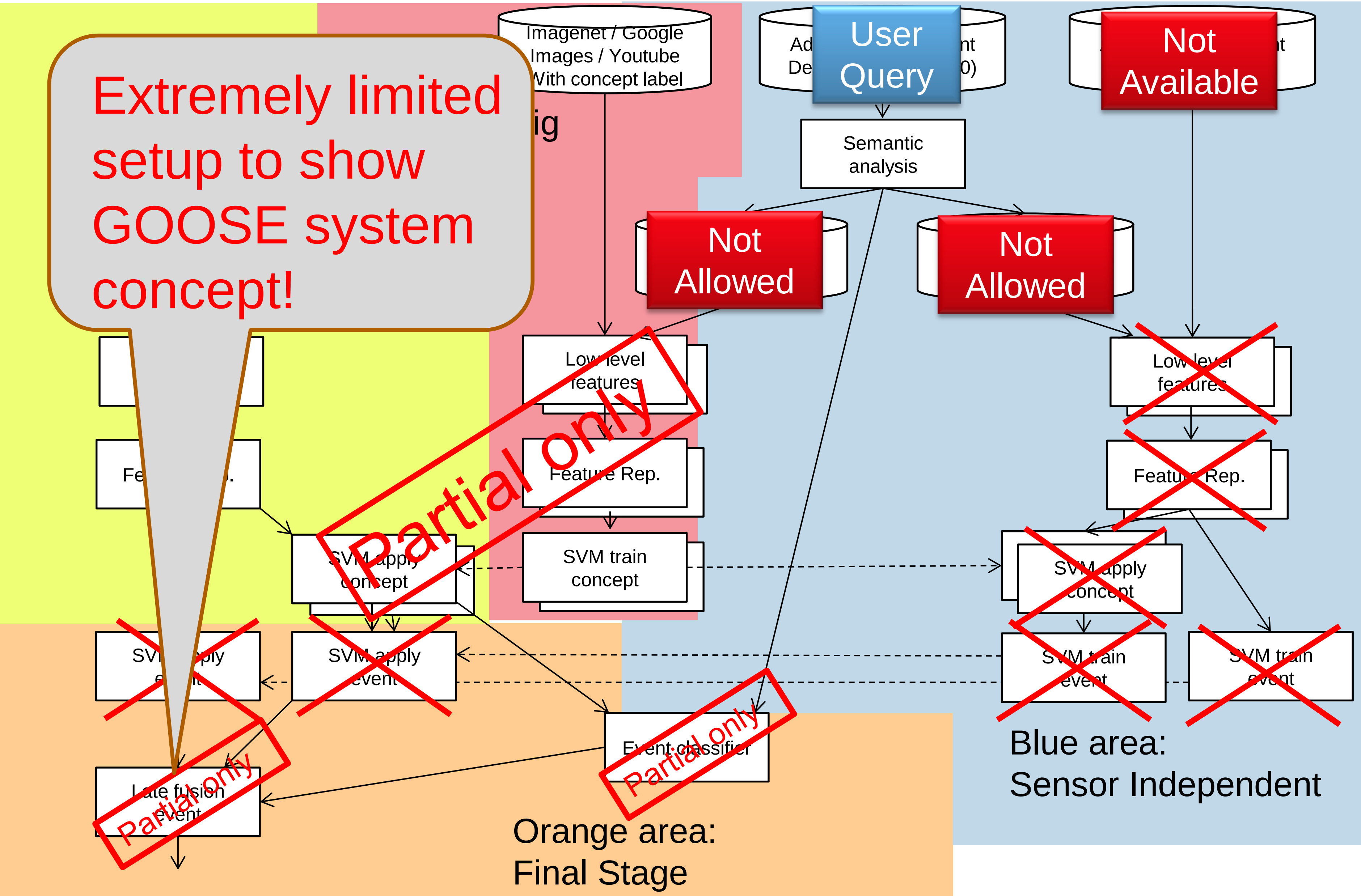
Inspired by MED 2012
CMU SESAME ECNU
BBNVISER SRIAURORA
MediaMill AXES Tokyo GENIE IBM



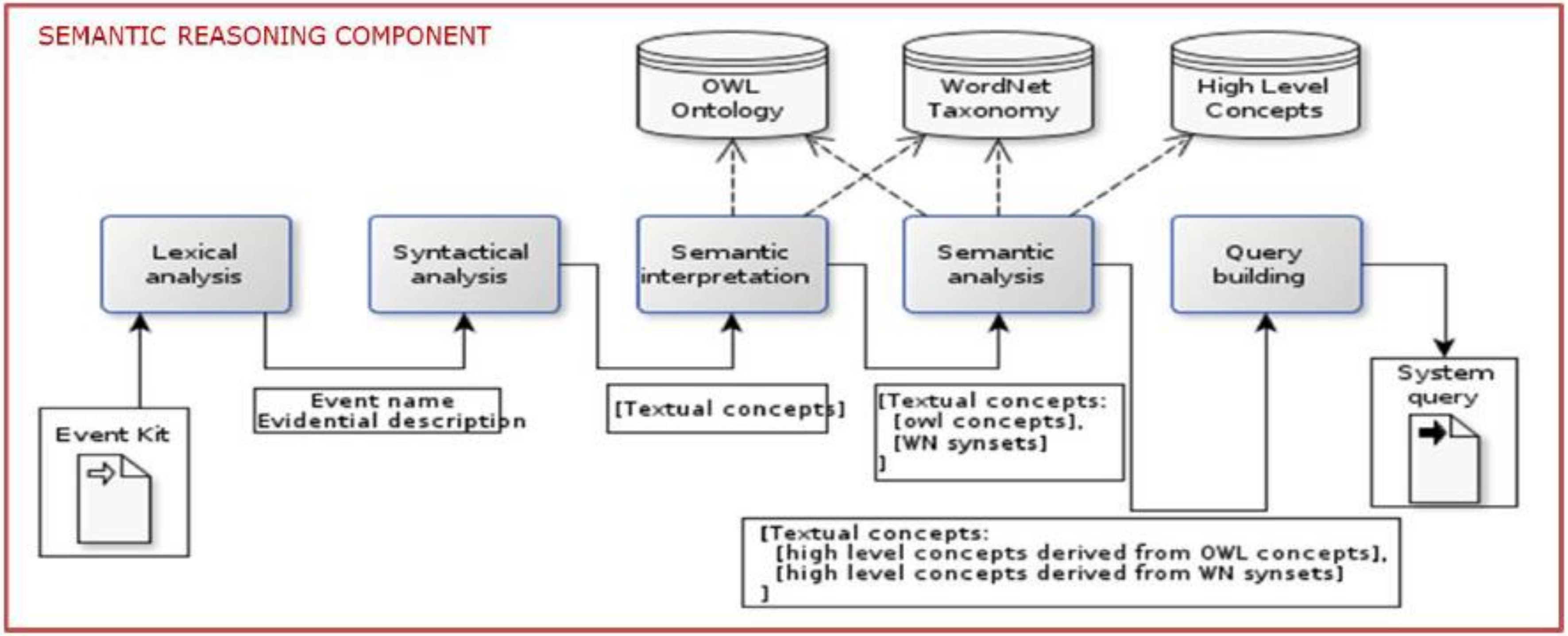
Applying TRECVID Pipeline to general GOOSE concept



Current MED limiting GOOSE: no online download!



Semantic Analysis flow



Semantic analysis – example

Win a race without a vehicle

Event Name		Winning a race without a vehicle
Evidential Description	scene	outdoors (park, field, track, road, or stadium) or indoors (indoor track, pool, or large gymnasium)
	objects /people	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
	activities	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
	audio	onlookers cheering, verbal or other indication of starting the race (yelling "Go!", gun shooting), narration of the race (speaking through a microphone)

Event Name

Evidential Description

Nouns

Verbs

Negations

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), circustent (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

Feature	Vocabulary size	Spatial tiling	Histogram size
SIFT	300	Spatial pyramid: 1x1 + 2x2 + 4x4	6300
Opp. SIFT	300	Spatial pyramid: 1x1 + 2x2	1500
LBP	300	2x2	1200
STIP	300	3x3	2700
MFCC	300	N / A	300

Unexpected results:

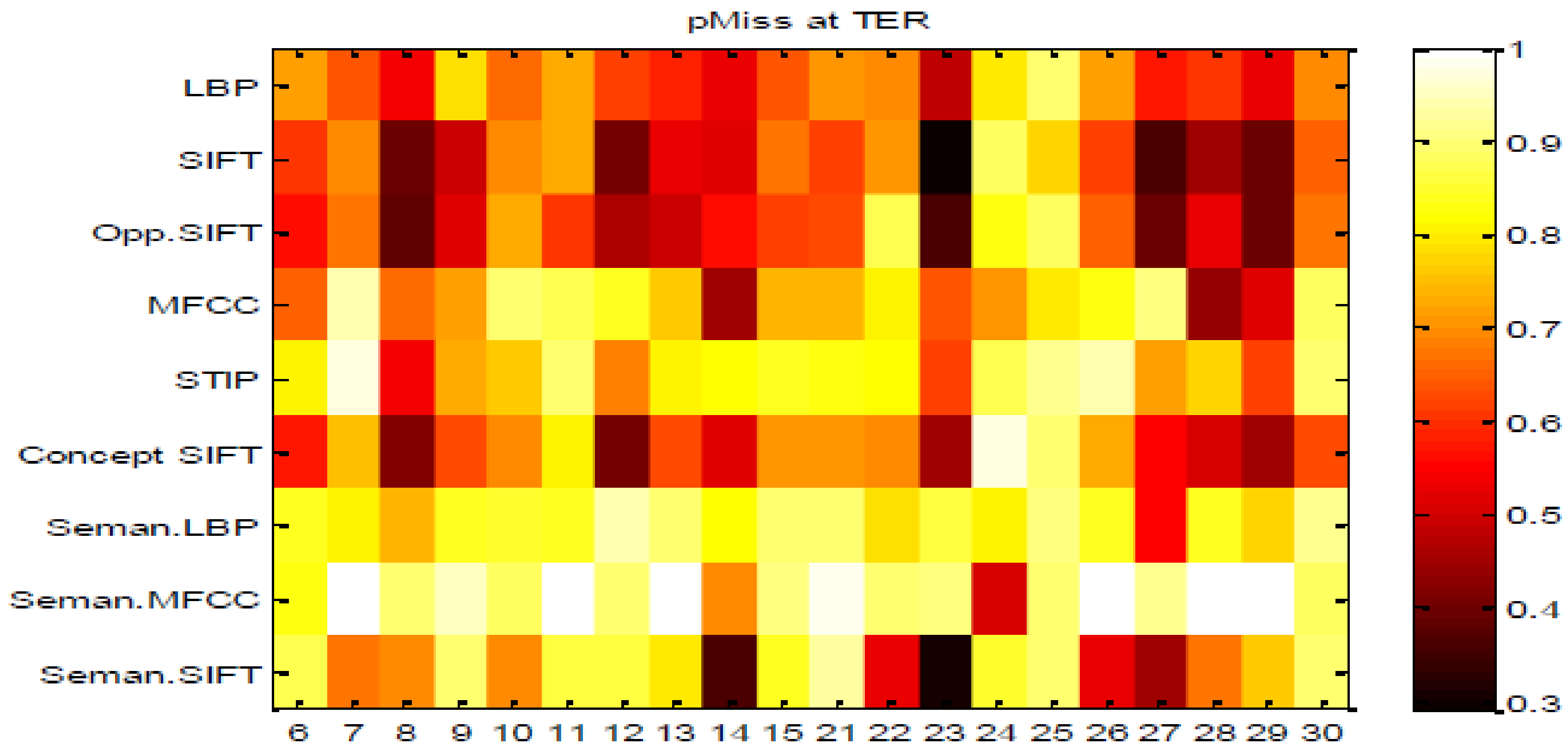
- Vocabulary size of 300 outperformed 100, 1000 and 3000
- VLAD (in combination with PCA) didn’t improve performance

Performance different features on training data

Note: MED Evaluation provided 2x3 numbers only!

	Pre specified (PS)		Ad hoc (AH)	
Method	pMiss @ TER (%)	pFA @ TER (%)	pMiss @ TER (%)	pFA @ TER (%)
Semantic: MFCC	90.2	7.3	91.0	7.2
Semantic: LBP	85.6	6.5	92.3	7.5
Semantic: SIFT	79.9	6.2	89.2	7.1
Feature: STIP	79.3	6.4	78.5	6.3
Fusion: 3 semantic (i 0 Ex	78.0	6.1	87.1	7.4
Feature: MFCC	74.0	6.0	74.3	6.0
Feature: LBP	65.7	5.2	63.6	5.1
Feature: Opponent-SIFT	59.1	4.7	55.8	4.5
Feature: SIFT Visual only	57.6	4.5	55.4	4.4
Fusion: 5 features	48.5	3.7	46.0	3.4
Fusion: 5 features + 3 semantic FullSys	47.8	4.0	46.4	3.4
Feature: CC D-STIP 1FPS *(not used)*	68.4	±5.5	---	---
Concept: SVM SIFT *(not used)*	63.7	5.1	---	---
Fusion: 5 features trained using on-the-fly downloaded video *(not allowed)*	---	---	63.0	5.1

Performance features versus events



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