Meanwhile at TRECVID

Progress in video concept and event search

Cees Snoek

University of Amsterdam

The Netherlands

Euvision Technologies
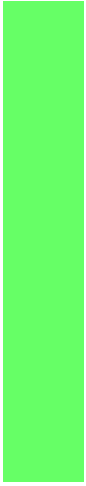
The Netherlands

NIST TRECVID benchmark

- Promote progress in video retrieval research
  - Provide common dataset
  - Challenging tasks
  - Independent evaluation protocol
  - Forum for researchers to compare results

http://trecvid.nist.gov/
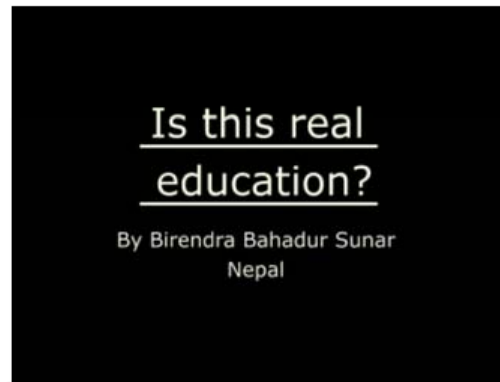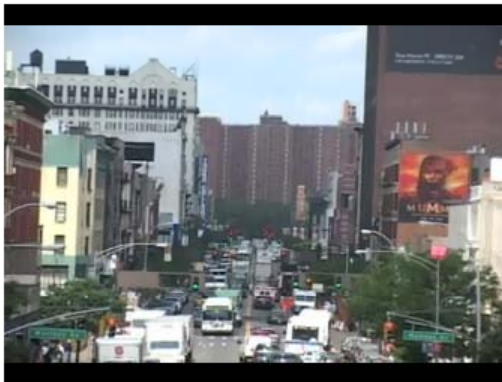
Challenging tasks

- Shot boundary detection

- Copy detection

- **Concept detection**

- **Multimedia event detection**

- Search

- …

CONCEPT DETECTION

More than broadcast news...

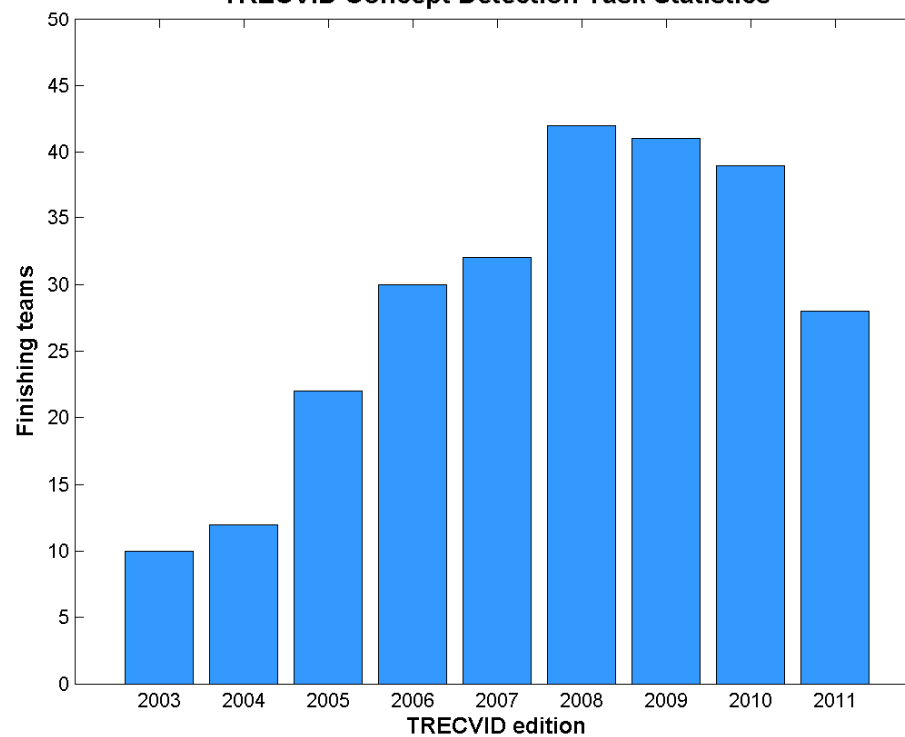- 600h of Internet Archive web video since 2010

De facto evaluation standard

Concept examples

- Aircraft
- Beach

- Note the variety in visual appearance

- Mountain
- People marching
- Police/Security
- Flowe

MediaMill concept detection

Software available for download at http://colordescriptors.com

Spatio-temporal sampling	Visual feature extraction	Codebook transform	Kernel-based learning

Are we making progress?

State-of-the-Art in Video-Concept Search
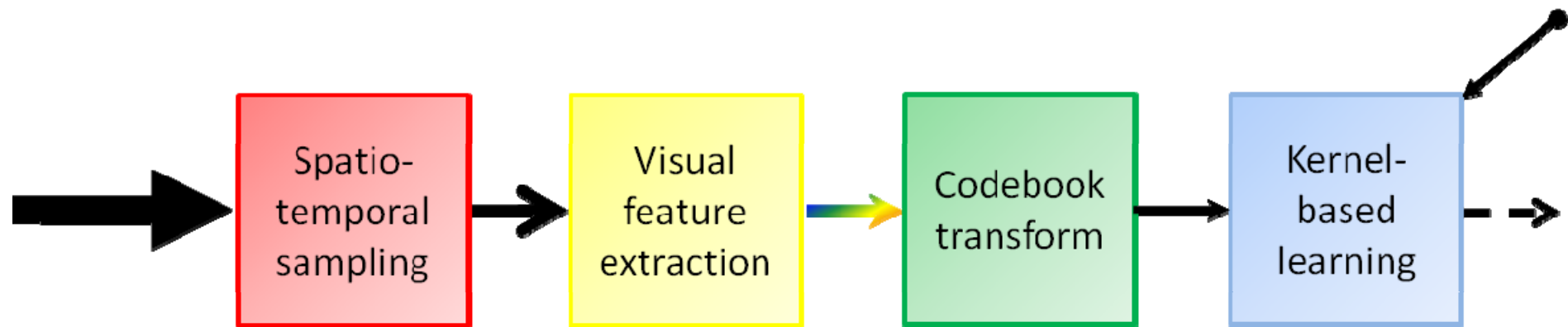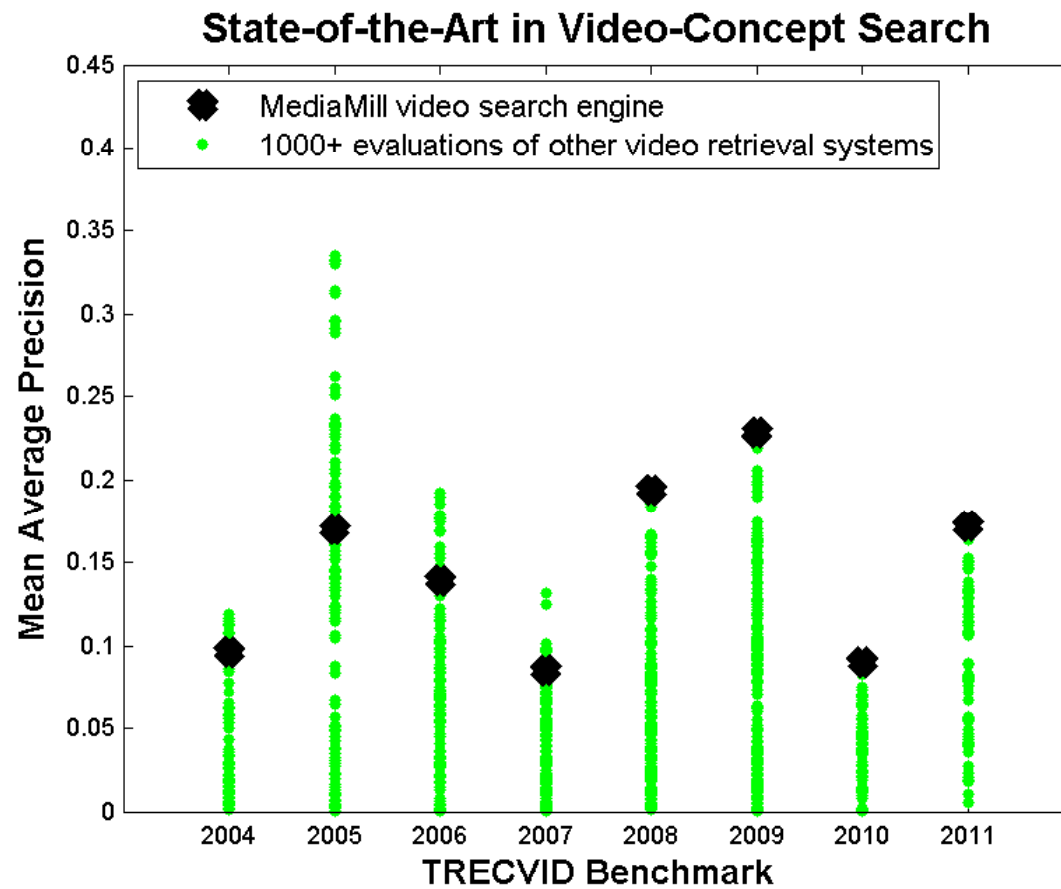
• >1000 other systems

Community myths or facts?

- Chua et al., ACM Multimedia 2007

  – Video search is practically solved and progress has only been incremental

- Yang and Hauptmann, ACM CIVR 2008
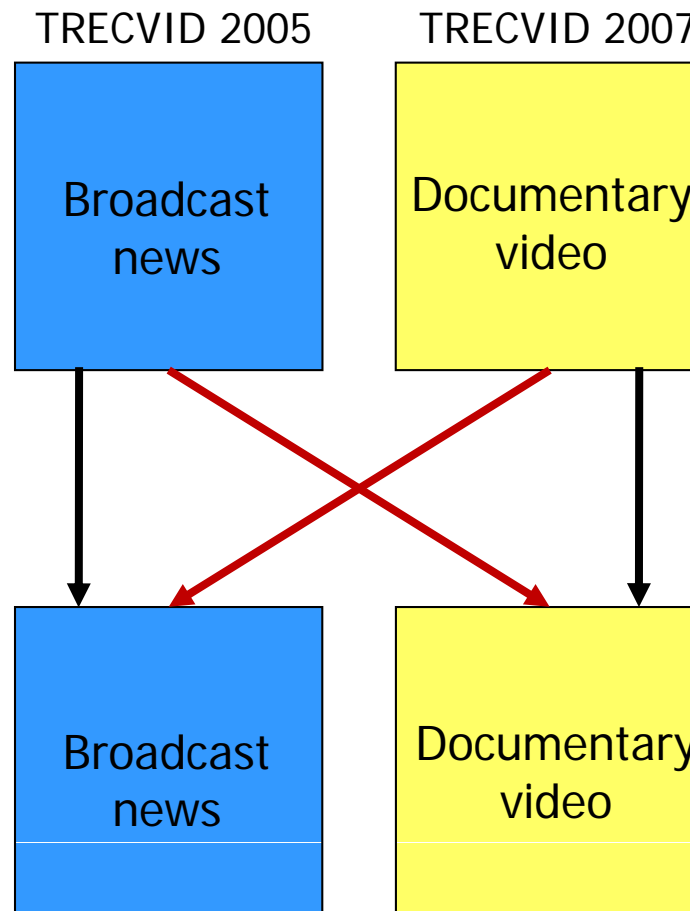
  – Current solutions are weak and generalize poorly

We have done an experiment

- Training

TRECVID 2005   TRECVID 2007

| Broadcast news | Documentary video |

- Testing

| Broadcast news | Documentary video |

Within domain

Cross domain

Performance doubled in just 3 years

• 36 concept detectors

– Even when using training
data of different origin

– Vocabulary still limited
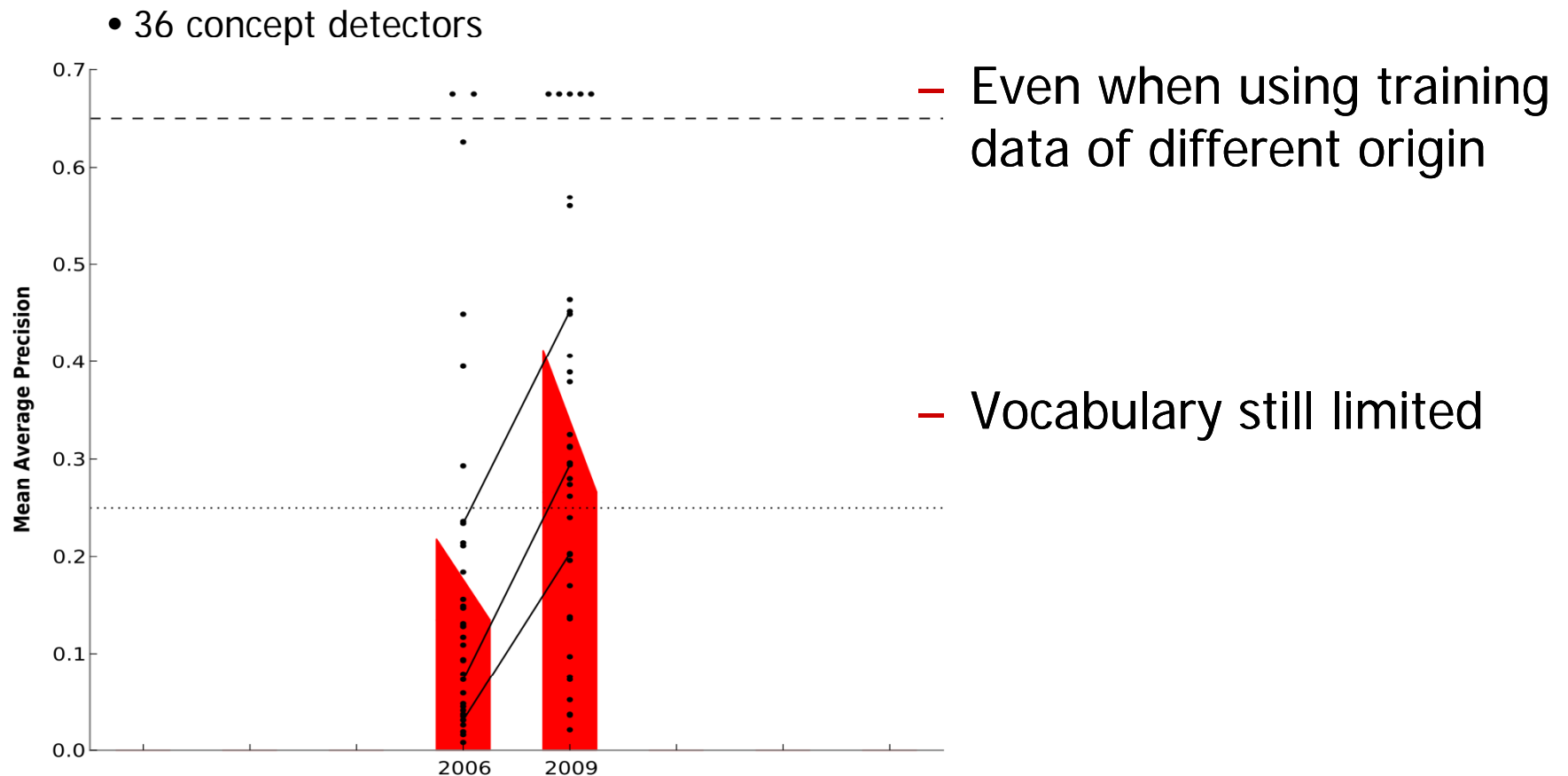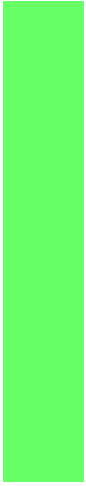
EVENT DETECTION

What defines an event?

Flash mob gathering   Changing a vehicle tire   Attempting a board trick

- …and how do we know it is really happening?

An event is a bag-of-features

Video

Feature vector

Event confidence
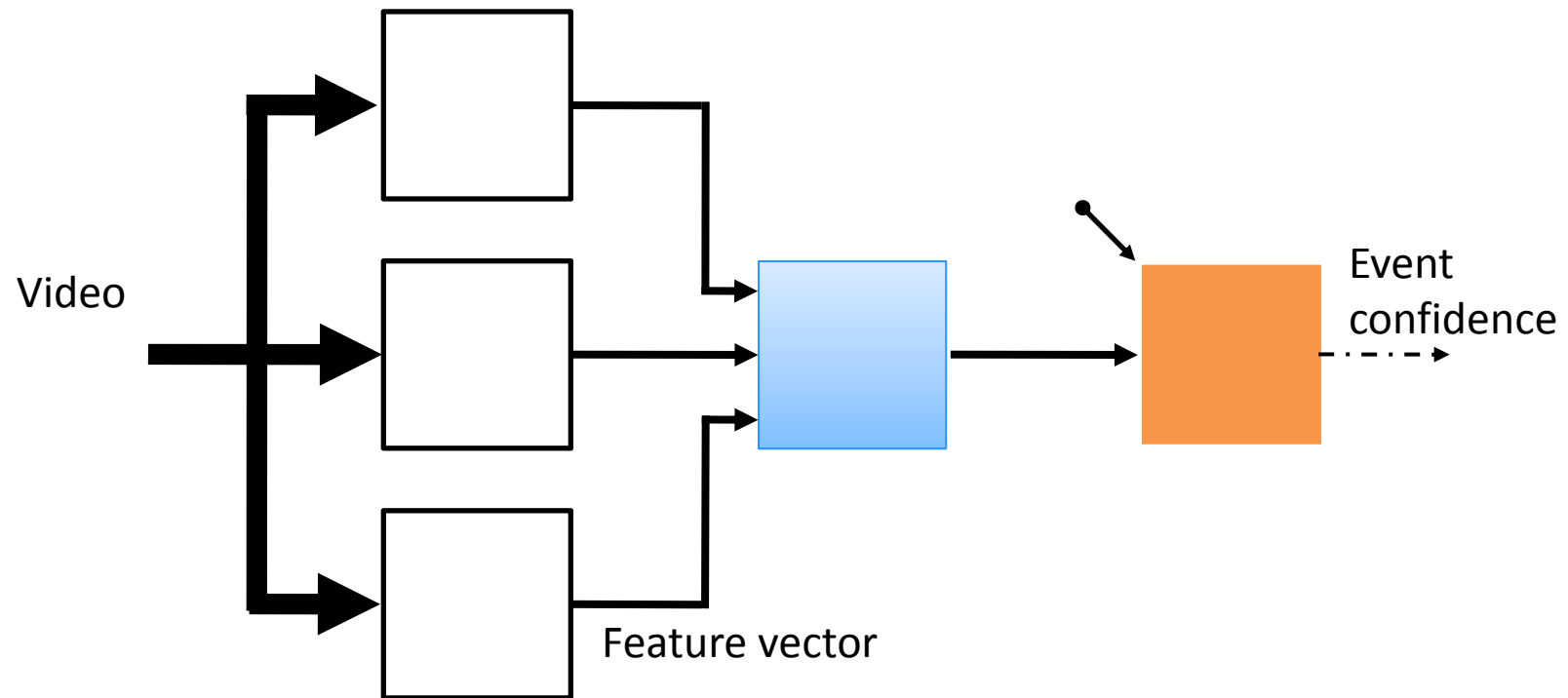
- Fusing multiple audio-visual features is effective
  - No notion whether the event is really found

An event is a bag-of-concepts
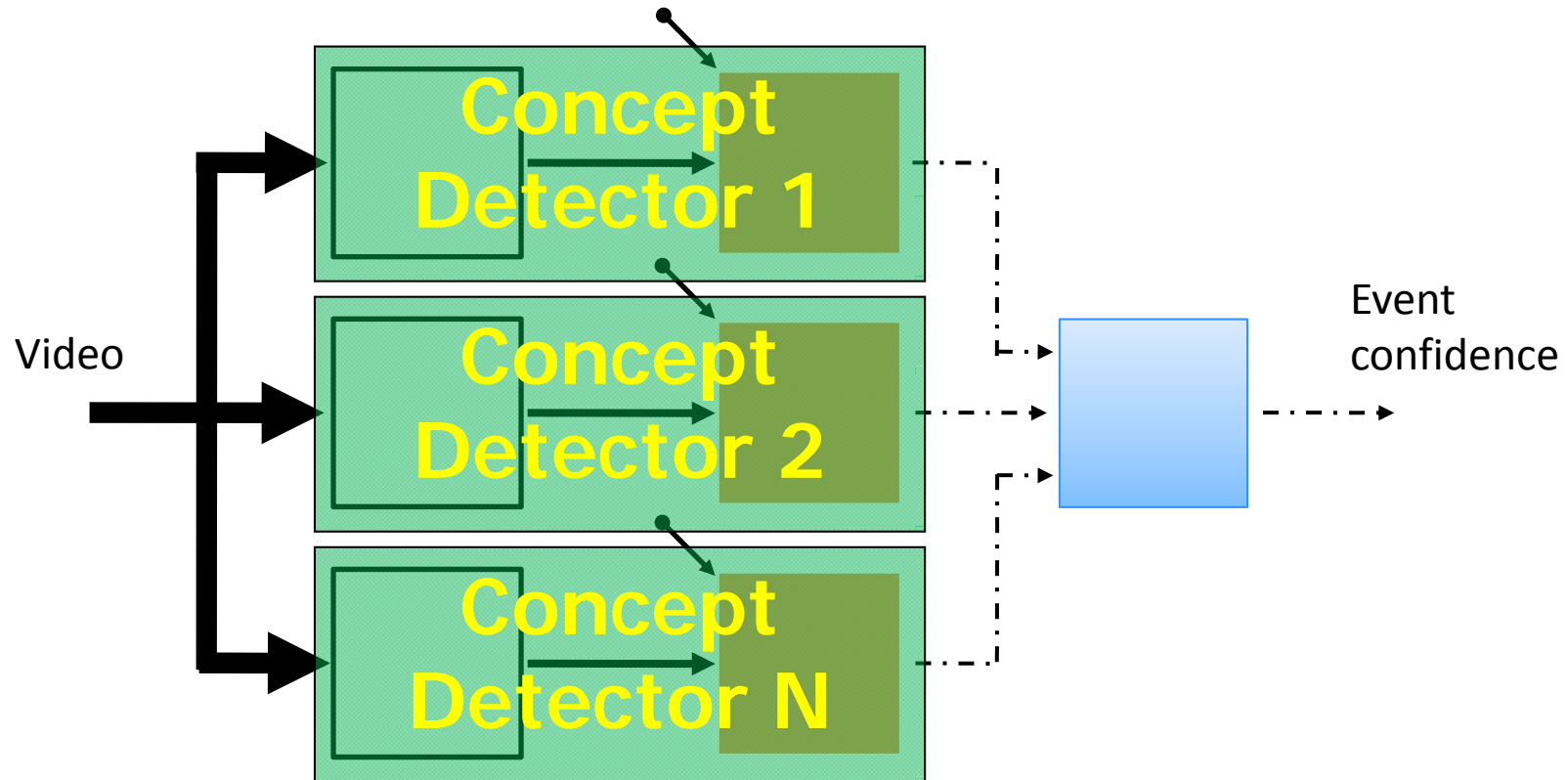
- Combine available detector scores in single vector
  - No notion what concept (ordering) is important for events

Horse, horse, horse?

Chair, chair, chair?

The event retrieval paradox

Multimedia Event Detection task
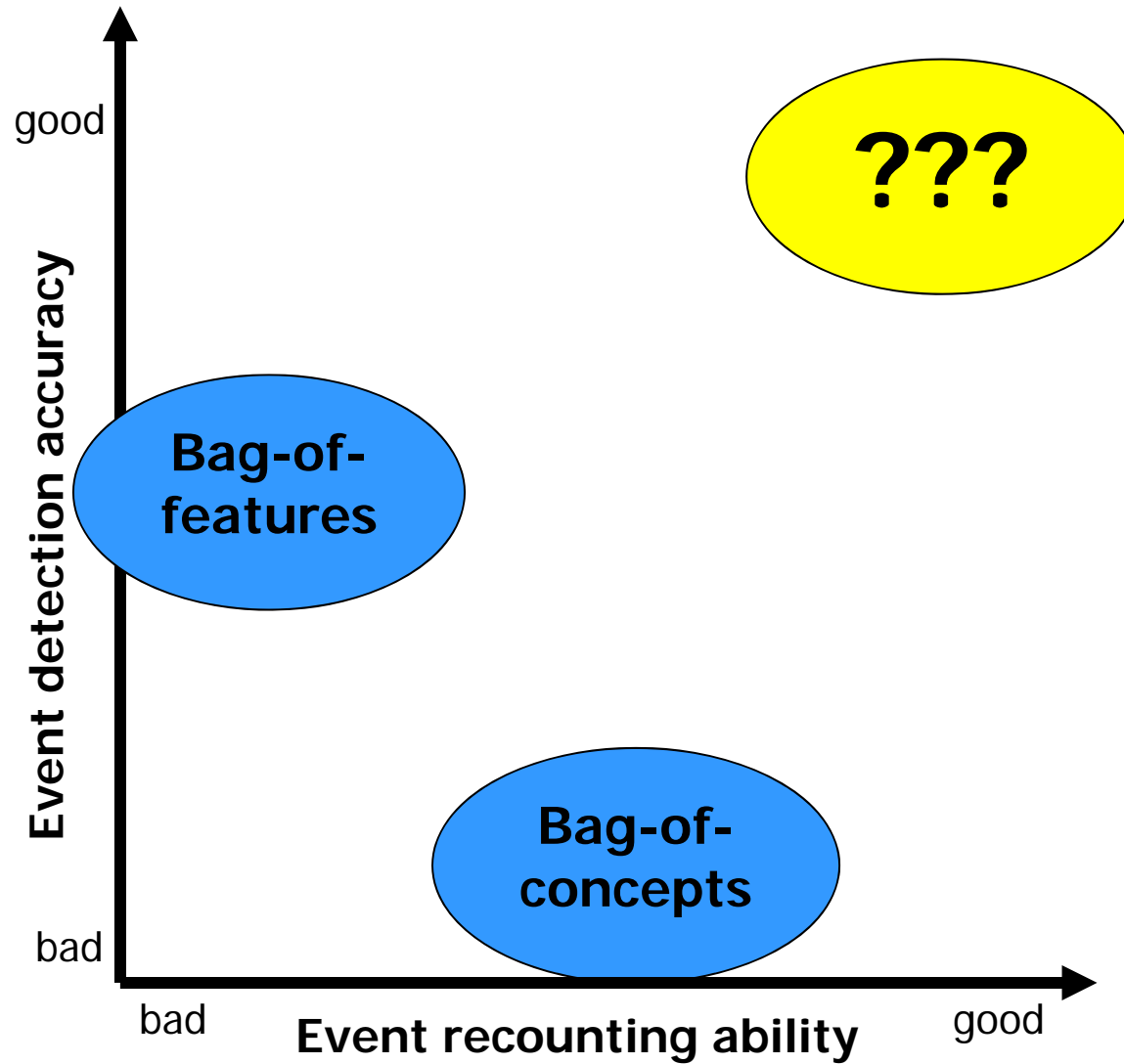
- Given an event specified by an
  - event kit (query) which consists of:
    - o definition,
    - o event explication,
    - o evidential description, and
    - o illustrative examples,

- Search collection for video's containing event

Example Event Kit

Event Name:

Working on a woodworking project

Definition:

One or more people fashion an object out of wood.

Textual Definition

Event Explication:

Woodworking is a popular hobby that involves crafting an object out of wood. Typical woodworking projects may range from creating large pieces of furniture to small decorative items or toys. The process for making objects out of wood can include cutting wood into smaller pieces …. (continues)

Expresses event domain specific knowledge to understand the event definition

Evidential Description:

scene: Often indoors in a workshop, garage, artificial lighting. Occasionally outdoors

objects/people: Woodworking tools (automatic or non-automatic saws, sander, knife), paint, stains, sawhorses, toolbox, safety goggles

activities: Cutting and shaping wood, attaching pieces of wood together, smoothing/sanding wood

audio: power tool sounds; hand tool sounds (hammer, saw, etc.); narration of process

Textual listing of attributes that are often associated with the event

Exemplars:

HVC334271.mp4, HVC393428.mp4, HVC875424.mp4, etc.

Specific clips from the "Event Kits" data set that are known to contain the event being defined.

Target User:

An *Internet information analyst* or *experienced Internet searcher* with event-specialized knowledge.

Internet video collections

Collection Name	Designated Uses	Target sizes	Annotation
Pilot	<u>2010</u> Development collection Test collection	1,723 clips 1,742 clips (100 hours)	Clip content annotation for both sets
Development (DEV)	<u>2011</u> Split into two subsets: (1) Transparent (DEV-T) (2) Opaque (DEV-O) <u>2012-2015</u> (1) and (2) merged to a single training collection	44K clips, (~ 1400 hours)	<u>For MED '11:</u> Clip content annotation for the transparent subset <u>After MED '11:</u> Clip content annotation for the opaque subset
Progress	<u>2012-2015</u>: test collection	120K clips, 4000 hrs	No clip content annotation
Novel 1	<u>2014</u>: test collection	120K clips, 4000 hrs.	No clip content annotation
Novel 2	<u>2015</u>: test collection	120K clips, 4000 hrs.	No clip content annotation

The TRECVID MED '11 events

Training Events

Process-Observed Events

Attempting a board trick
Feeding an animal
Landing a fish
Working on a woodworking project

Life Events

Wedding ceremony

Testing Events

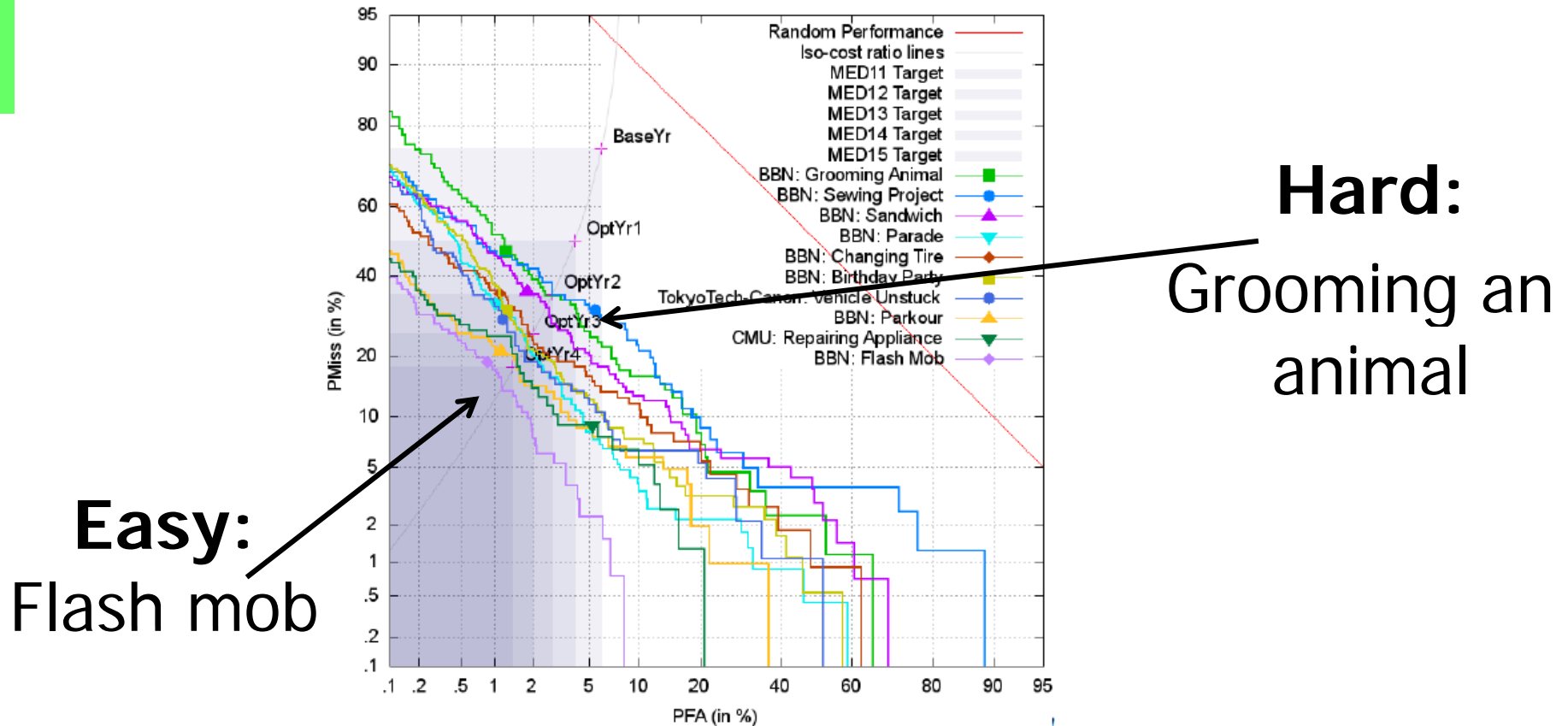**Process-Observed Events**

Changing a vehicle tire
Getting a vehicle unstuck
Grooming an animal
Making a sandwich
Parkour
Repairing an appliance
Working on a sewing project

Life Events

Birthday party
Flash mob gathering
Parade

2011 event detection results

All systems rely predominantly on bag of features,
no notion whether event really happened

Conclusion

- TRECVID promotes and delivers progress
    - Concept detection performance doubled
    - Event detection performance to follow?

Thank you

- dr. Cees Snoek

  www.ceessnoek.info

  cgmsnoek@uva.nl

  twitter.com/cgmsnoek