

# Video Semantics and the Sensor Web

Alan F. Smeaton

**CLARITY: Centre for Sensor Web Technologies**  
**Dublin City University**  
**Alan.Smeaton@DCU.ie**

# Talk Overview

- o CLARITY, what it is
- o Proposition of the sensor web
- o Some CLARITY examples of sensor web in different domains ... lifelogging, tennis, cycling, environment
- o TRECVID ... semantics from video
- o Some parallels, some relationships, some challenges



# Video, Semantics and the Sensor Web

Alan F. Smeaton

CLARITY: Centre for Sensor Web Technologies  
Dublin City University  
Alan.Smeaton@DCU.ie

# CLARITY



## **CLARITY: Centre for Sensor Web Technologies (2008 – 2013)**

National, multi-site research centre funded by Science Foundation Ireland & industry;

5 Year Programme with strong Industry collaboration;

Significant presence in Ireland's research infrastructure, very collaborative

## **Earlier work proved the 'Sensor Web' Proposition - key developments in crucial Sensor Web technologies**

- Wearable technologies
- Environmental monitoring
- Personalized media

## **Wide range of domains**

- Physiotherapy, Sports Science, hardware, networking, engineering, clinicians, neuroscience, new and traditional media
- Significant international collaborations (EU, US)

# CLARITY: Multi-disciplines

## Principal Investigators (PIs)

Prof. Barry Smyth	- <i>Personalization, recommender systems, mobile computing</i>
Prof. Alan Smeaton	- <i>Content-based information retrieval</i>
Prof. Dermot Diamond	- <i>Materials research, wearable sensors</i>
Prof. Noel O'Connor	- <i>Audio-visual analysis, multi-modal information processing</i>
Prof. Gregory O'Hare	- <i>Ubiquitous computing, multi-agent systems</i>

## Associate PIs

Prof. Paddy Nixon	- <i>Pervasive computing, middleware, security, trust, privacy</i>
Prof. Niall Moyna	- <i>Sports Science, wearable sensing</i>
Dr. Simon Dobson	- <i>Middleware, pervasive computing</i>
Dr. Cian O'Mathuna	- <i>Sensor devices, energy-aware hardware</i>
Dr. Brian Caulfield	- <i>Physiotherapy, therapeutic gaming, wearable sensors</i>

## Collaborators

Chris Bleakley (UCD), Conor Brennan (DCU), Rem Collier (UCD), Brian Corcoran (DCU), Cathal Gurrin (DCU), Neil Hurley (UCD), Lorraine McGinty (UCD), Kieran Moran (DCU), Kieran Nolan (DCU), Brendan O'Flynn (TNI), Donal O'Gorman (DCU), Brett Paull (DCU), Emanuel Popovici (TNI), Aaron Quigley (UCD), Mark Roantree (DCU)

# The Sensor Web

## The 'Old' Web

## The Social Web

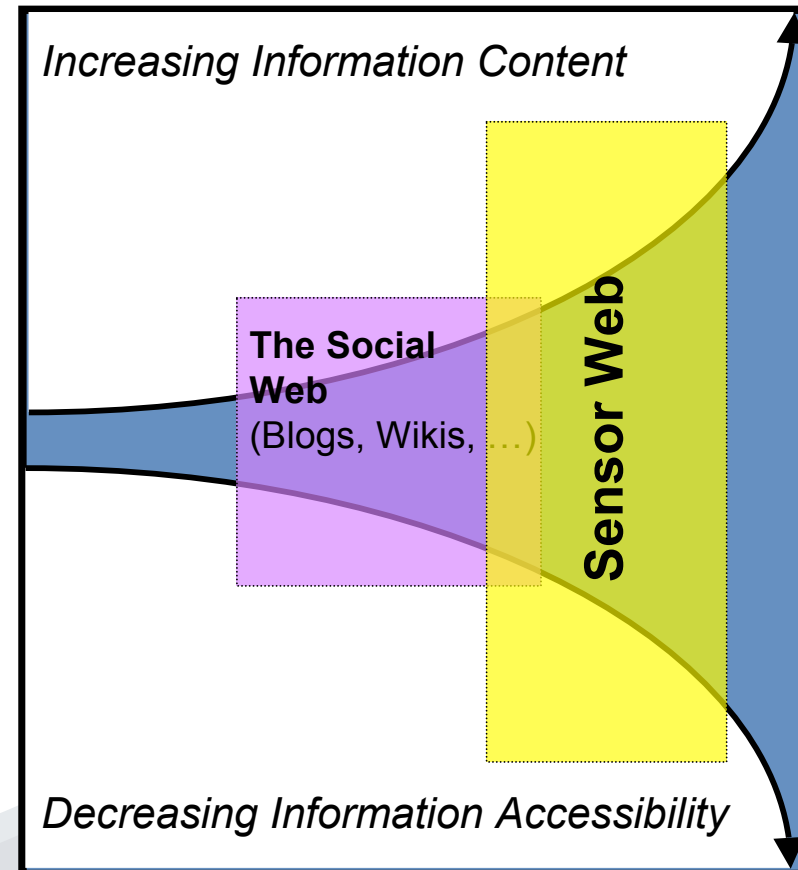
Dynamic user generated content (UGC) - conventional search tools unable to cope

## The Semantic Web

In here somewhere ...

## The Sensor Web

Increasing availability of cheap, robust, deployable sensors as ubiquitous information sources  
Key is that they are networked, mostly wireless, hence global, and integrated - WSNs  
Dynamic and reactive but noisy, and unstructured data-streams



# CLARITY Applications

## Demonstrator Projects ... unifying spaces for research threads

- o Multi-model sensing for sports .. Tennis, Hockey, Cycling, GAA, ...
- o Environmental Sensing Platforms .. water quality, air quality
- o Personal Environmental Impact Monitoring .. 'Karbon' footprint
- o Ambient Assisted Living

**We will look at some of these  
as examples of sensors in action**

**For each, we will ask ...  
“what do people want”**

**But first ... lifelogging**

# Sensor App 1: Lifelogging

**Lifelogging is digitally recording daily life**

**Sometimes for a reason**

- **Work e.g. security personnel, medical staff, etc.**
- **Personal e.g. diaries, etc.**

**Sometimes for posterity**

- **Recording vacations, family gatherings, social occasions**

**Sometimes because we can**

- **and we're not yet sure what we'll do with it e.g. MyLifeBits**

# Visual Lifelogging Devices

Much past research focus on miniaturising hardware and increasing battery-life + storage e.g. visual lifelogging domain



Steve Mann. Wearable computing: a first step toward personal imaging. *Computer*, 30:25–32, Feb 1997.

TIMELINE



Tano *et. al.* University of Electro-Communications, Tokyo, Japan



Microsoft Research  
SenseCam



# SC data: c.3.5M SenseCam Images

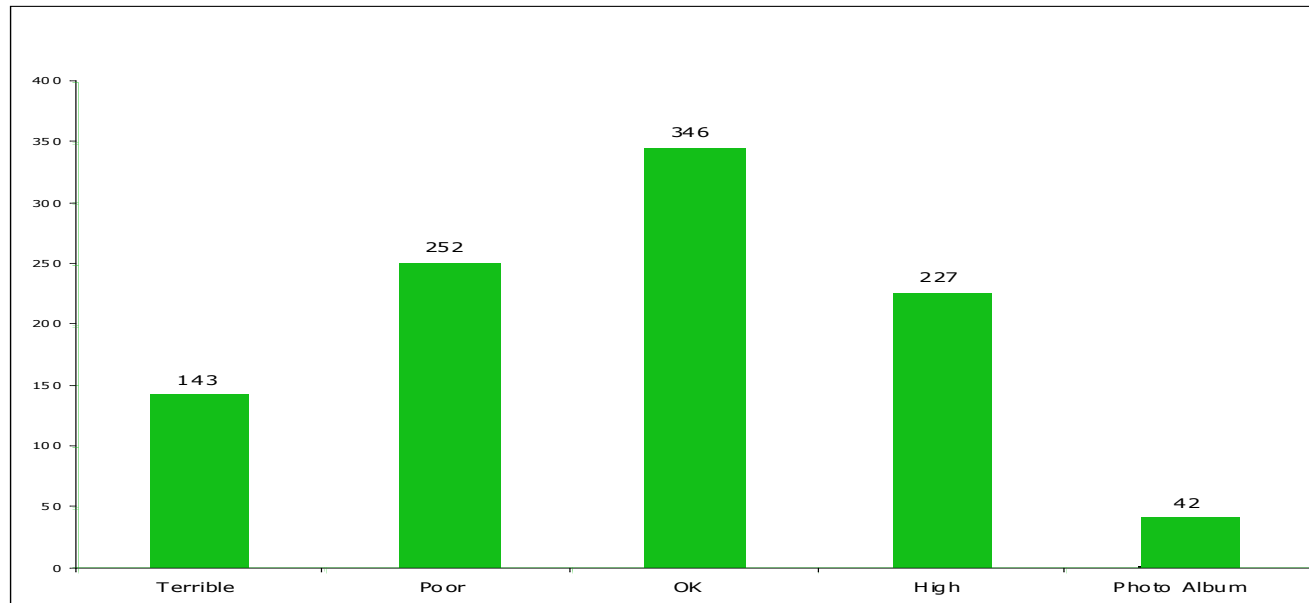
- **One user wears SC for over 3 years, all day**
  - Each with GPS position
- **Experiences:**
  - Most people don't notice camera
  - Those that do always remember !
  - Most people don't mind the camera
  - Have been spotted/greeted by people who have heard about the 'guy with the camera'



# SenseCam Lessons

- Event browsing is key
  - Too many photos to browse, need summary of semantic events and then ‘drill down’ to view event detail if required
- Stop events, (like work desk and driving) should be identified and hidden.
- ‘Total Recall’, little sign of ‘Event Decay’
  - Remembers nearly every (non stop-) event when presented again
  - Experiment underway w/ U.Leeds exploring this
- Lessons from three years research and exposure suggest that lifelogging is of interest to many

# SC Image Quality



- 40% of images are of low quality
- Many “boring” images of mundane tasks

Despite this, we can have great (technological) fun ...

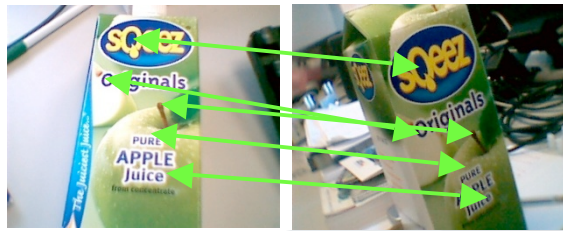
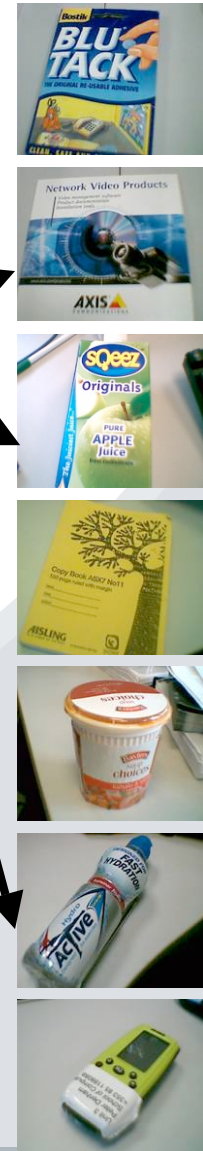
# Advanced Image Matching



SURF features are extracted

Each feature point casts a weighted vote for multiple database images

Votes are accumulated & the best match is found



# Setting Detection – Watching TV

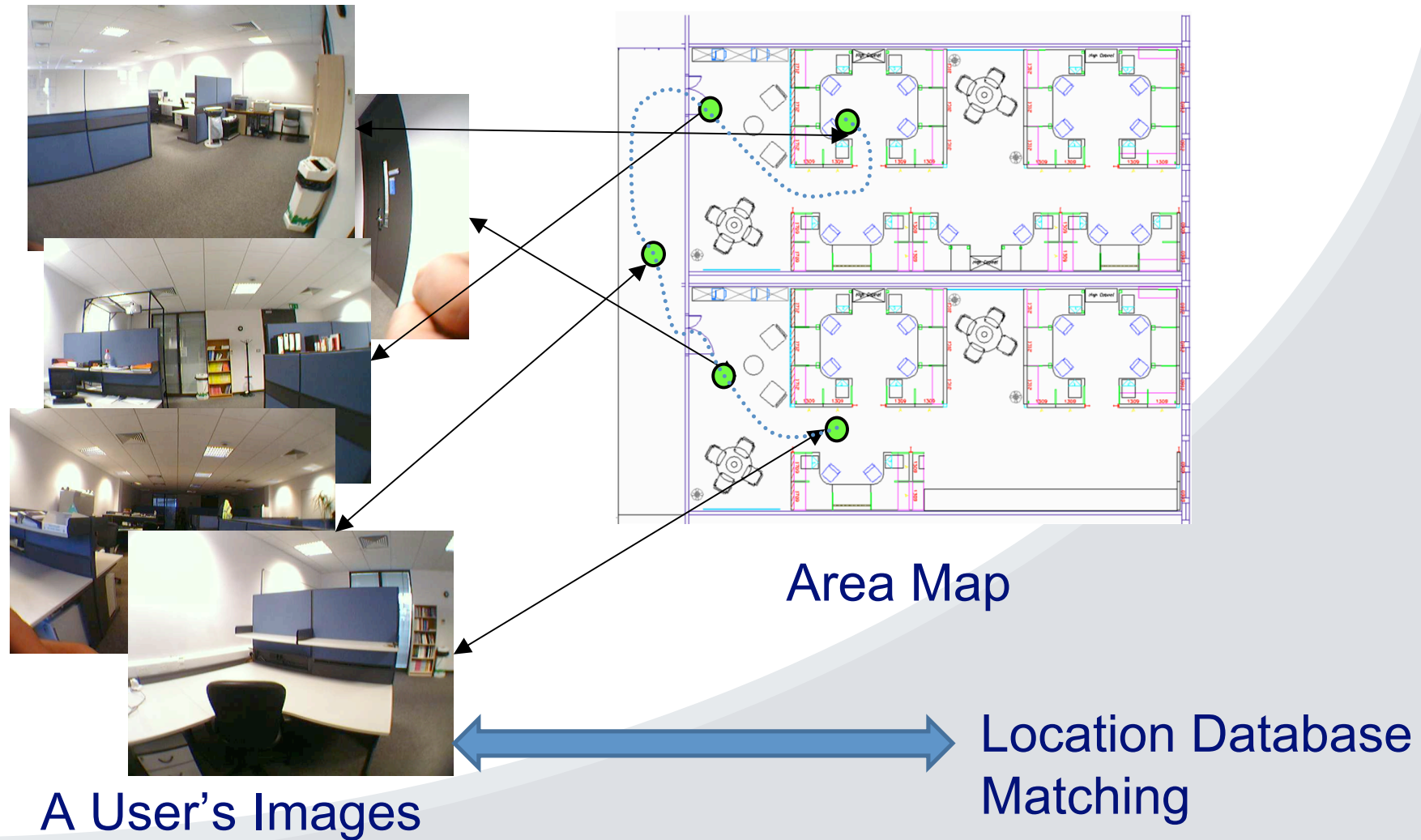




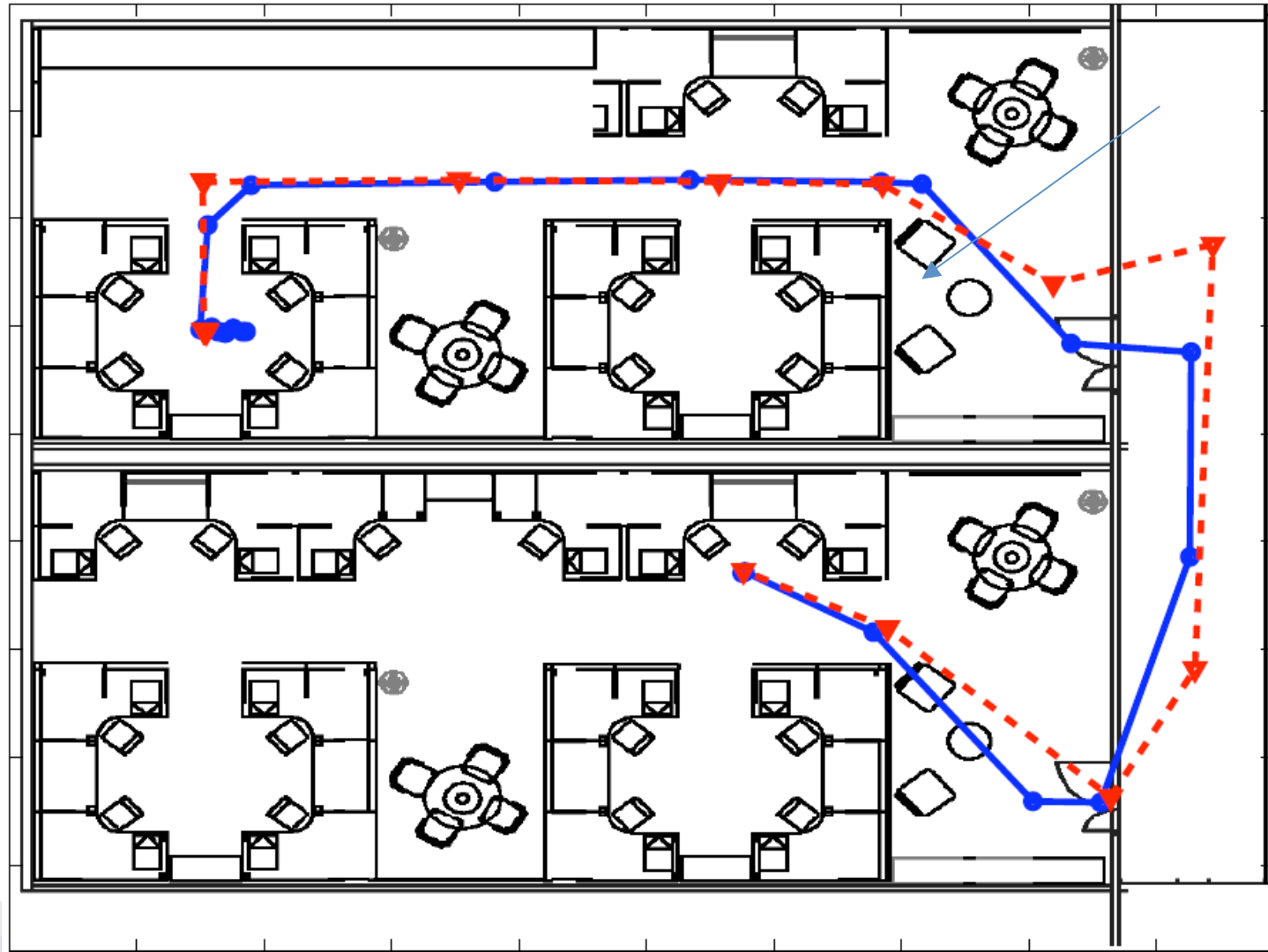
# Setting Detection – In the Park



# Trajectory Estimation



# Trajectory Estimation Results



Ground  
truth



# Activity Recognition

27 “concepts”

Outputs manually judged on ~95k images (5 users)



Vehicles External(46%)



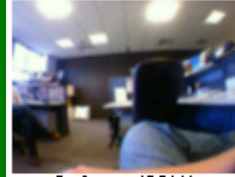
Road (47%)



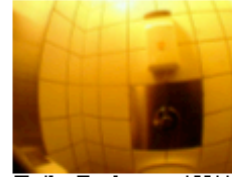
Steering wheel (72%)



Inside of vehicle (60%)



Indoors (82%)



Toilet/Bathroom (58%)



Door (69%)



Staircase (48%)



Outdoors (62%)



Buildings (59%)



Tree (63%)



View of Horizon (23%)



Grass (60%)



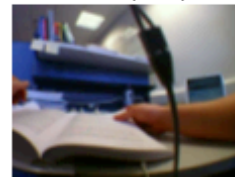
Sky (79%)



Vegetation (64%)



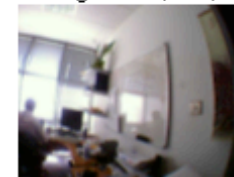
Screen (78%)



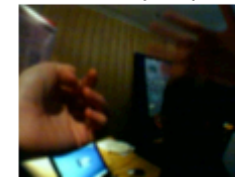
Reading (58%)



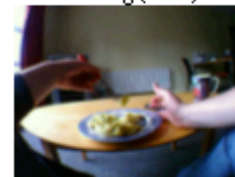
Meeting (34%)



Office (72%)



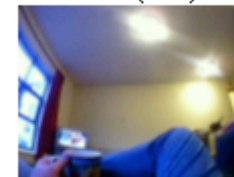
Presentation (29%)



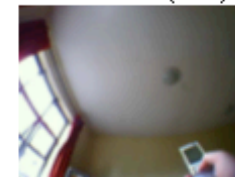
Food/eating (41%)



Hands (68%)



Holding cup (35%)



Holding phone (39%)



Faces (61%)



People (45%)



Shopping (75%)

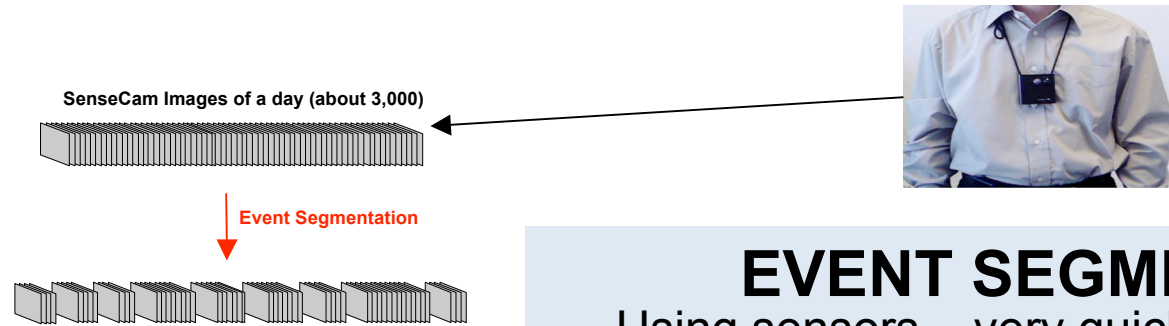
# But what do people want ?

Memory science says ... to effectively provide memory retrieval cues using SenseCam images and sensor readings, we need to automatically:

- *Group similar images into distinct semantic “events”*
- *Suggest more interesting/distinctive semantic events*
- *Associate related semantic events*
- *Provide additional retrieval cues from other sources*

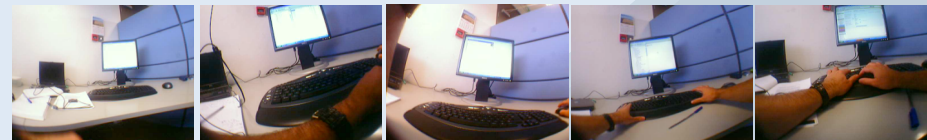
Called *cued recall*, trigger our own memories, not a memory replacement

# Daily Overview



## EVENT SEGMENTATION

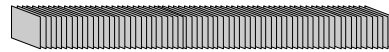
Using sensors – very quick & accurate



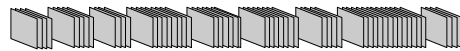
# Visual Search Facilities



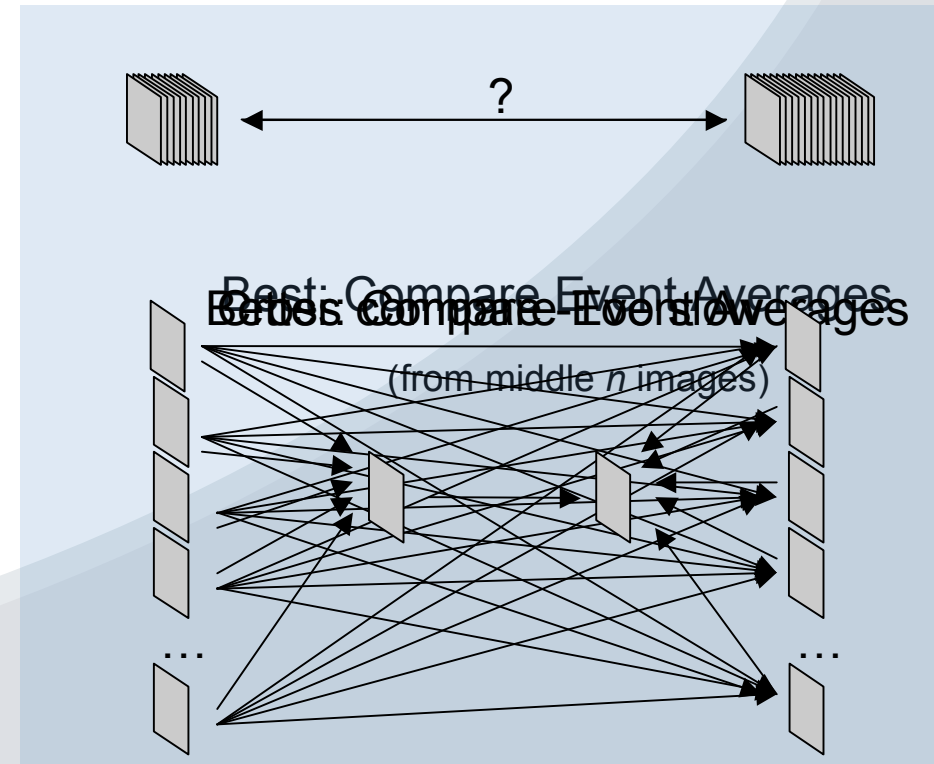
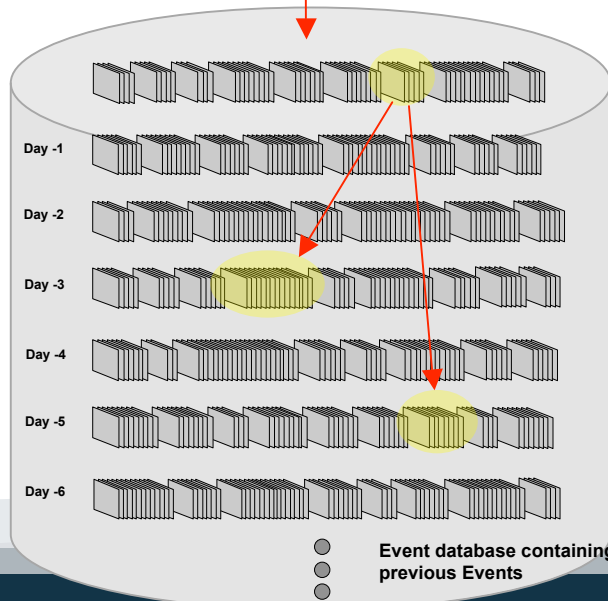
SenseCam Images of a day (about 3,000)



Event Segmentation



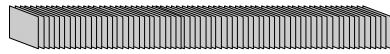
Event-Event Comparison  
within the Multi-day Event  
database



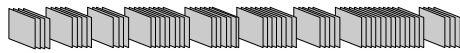
# Selecting Event “Keyframe”



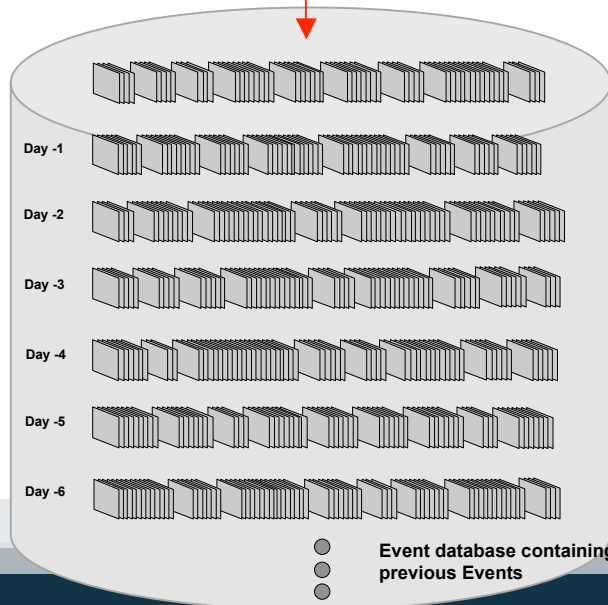
SenseCam Images of a day (about 3,000)



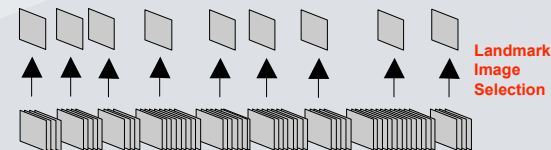
Event Segmentation



Event-Event Comparison  
within the Multi-day Event  
database

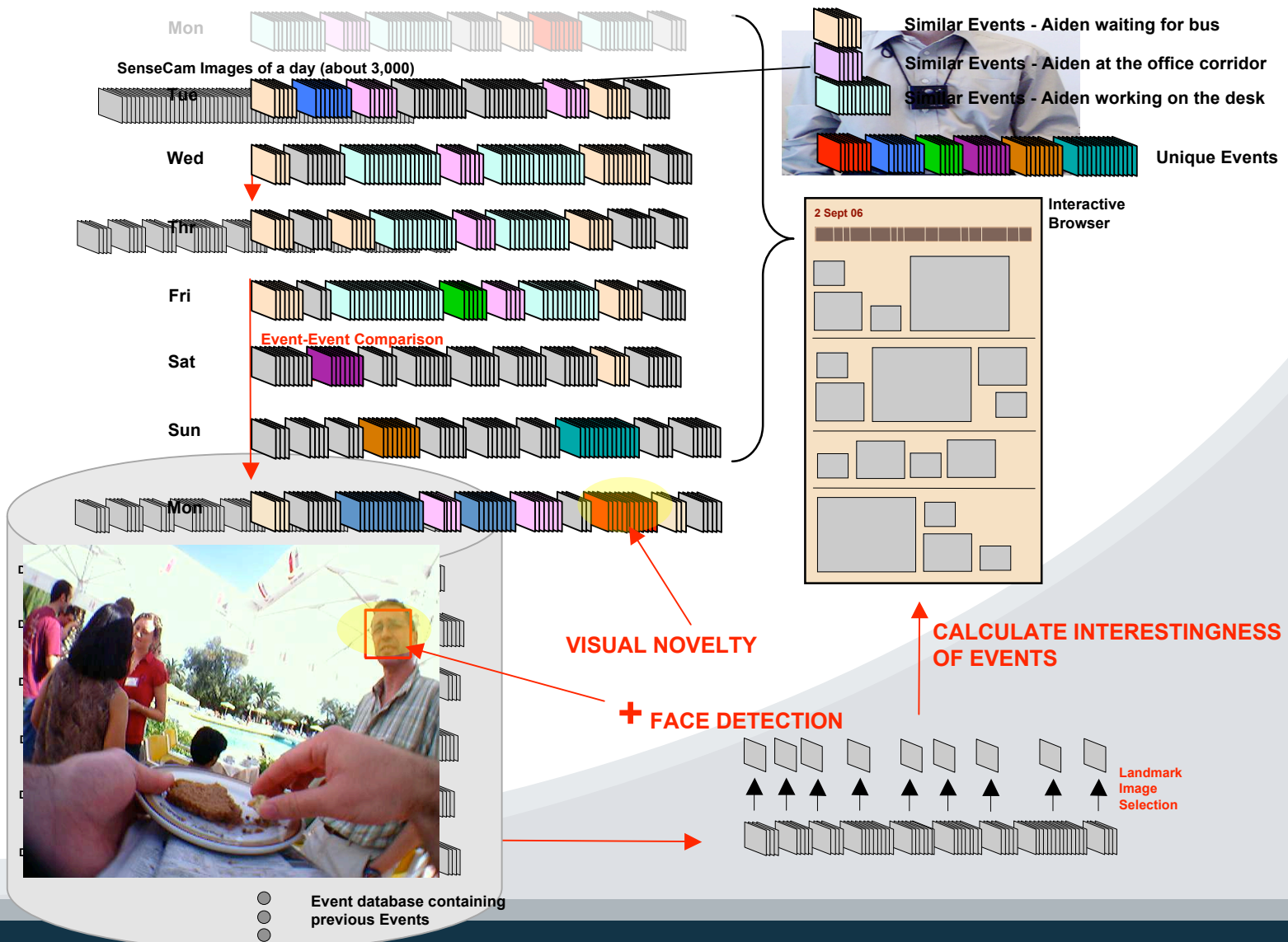


Best QUALITY  
image around  
MIDDLE of event





# Suggest Interesting Events



# Event augmentation

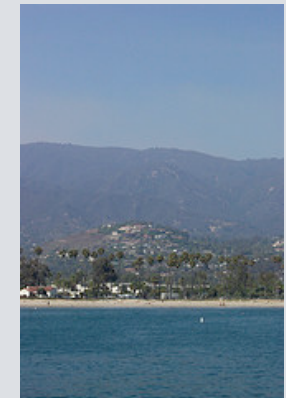
SenseCam picture at a pier in Santa Barbara

WITH GPS we can search for other pictures in the same location...



# Event augmentation – more cues

- We receive the following geotagged images...
- Then after some processing on text associated with these images we get many more images, and even YouTube videos
- Then do visual filtering to choose those for SC event augmentation





# Event augmentation

Operational from 6 web sources, tested and evaluated with users.

Bringing the threads together ... segmentation into semantic events, KF selection, event importance, event searching, and event augmentation ...

... a system to manage a lifelog



## CALENDAR

◀ MAY ▶ 2006 ▶

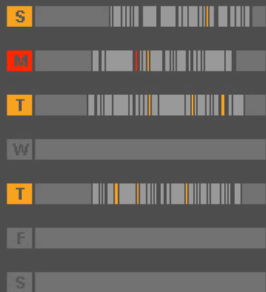
S	M	T	W	T	F	S
30	1	2	3	4	5	6
7	8	9	10	11	12	13
14	15	16	17	18	19	20
21	22	23	24	25	26	27
28	29	30	31	1	2	3
4	5	6	7	8	9	10

DURATION ▶

## CAPTION SEARCH

### WEEKLY SUMMARY

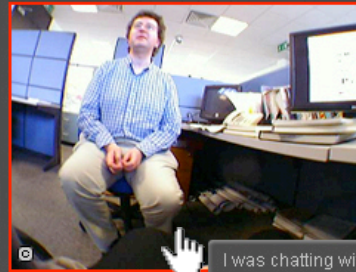
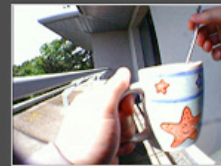
Selected day is shown below in the context of whole week. Move mouse cursor over to see other similar Events in the week



### 29 May 2006

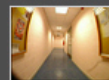
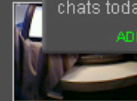
19  
EVENTS

Drag the slider bar to adjust the number of Important Events



I was chatting with Gareth on the conference in July. Quite a few chats today! ↻ x

ADD TO FAVE | FIND SIMILAR



MY ACCOUNT | SIGN OUT | ABOUT

## My FAVOURITE EVENTS

25 Favourite Events are shown below. Click on the photo to replay all photos within the Event.

| 1 | 2 | 3 |

Sort by: **TIME** | SIMILARITY | #PEOPLE



16:20 (Duration: 08m 43s)  
14 APR 2006 ▶



13:45 (Duration: 14m 05s)  
14 APR 2006 ▶



10:02 (Duration: 23m 56s)  
13 APR 2006 ▶



14:39 (Duration: 15m 30s)  
12 APR 2006 ▶



11:25 (Duration: 06m 21s)  
12 APR 2006 ▶



09:52 (Duration: 01m 03s)  
12 APR 2006 ▶

# Lifelogging Roadmap ...

Device miniaturisation

*Stories* from memories

Ambient replay

Autocueing for recall

Other sources

# Sensor App 2: Tennis

## Initial focus on tennis, indoor

Pairs or doubles, well-structured game, intense physical activity, non-contact sport

## Then move to team sport

Probably still non-contact, team tactics, outdoor, field hockey and GAA

## Wide definition of sensors

A/V, movement and location, physiology



# Body Sensors

Use off the shelf technology

## Polar Heart Rate Monitor



## BodyMedia SenseWear Armband

Galvanic Skin Response (GSR), heat flux, skin temperature, accelerometer



## Foster Miller vests

Respiration, body temperature, heart rate, GPS  
Robust, comfortable, re-usable





# Infrastructure



**Tennis Ireland, National Tennis Federation**  
**Dublin City University**

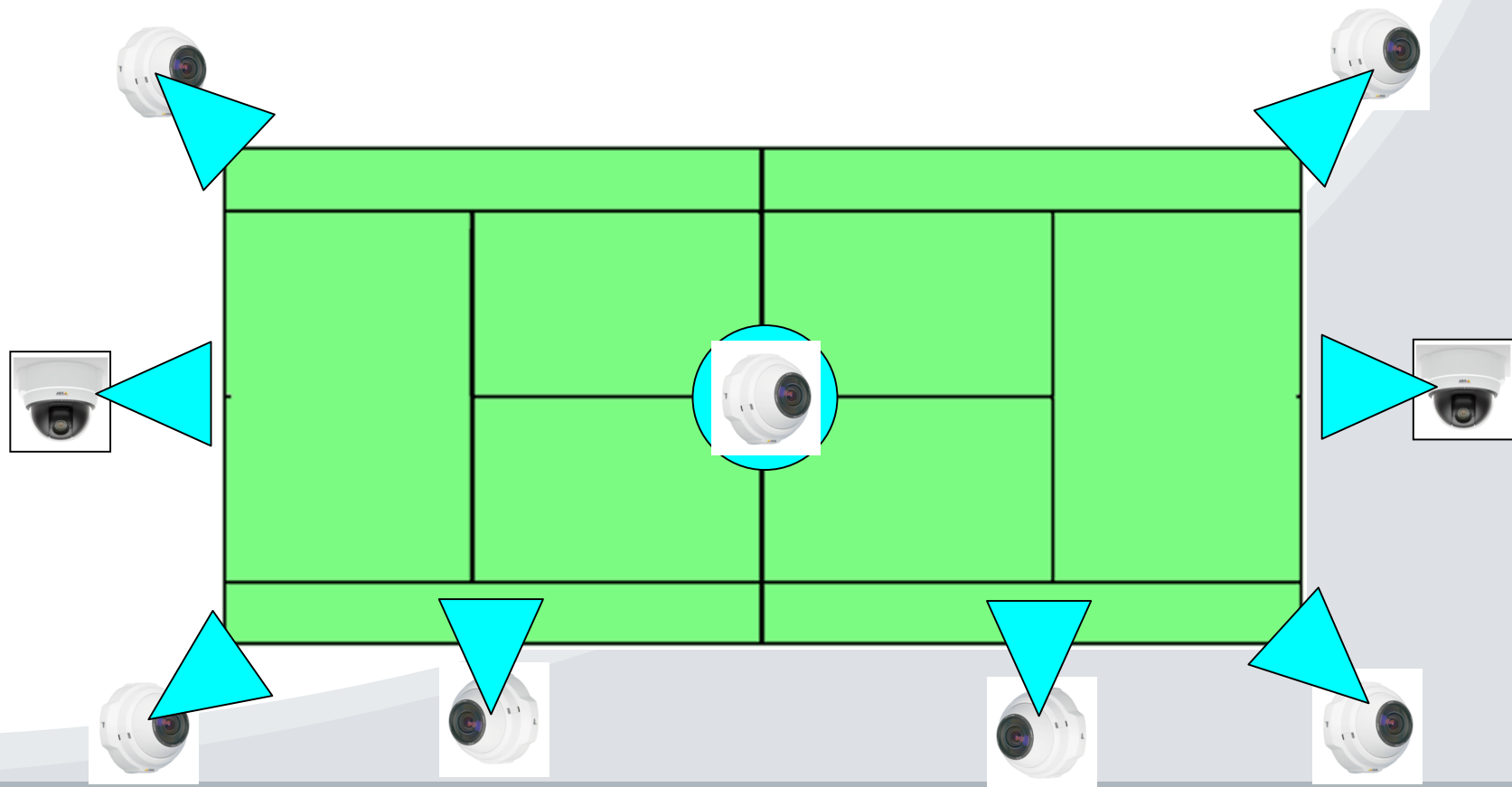
# Audio-Visual Network



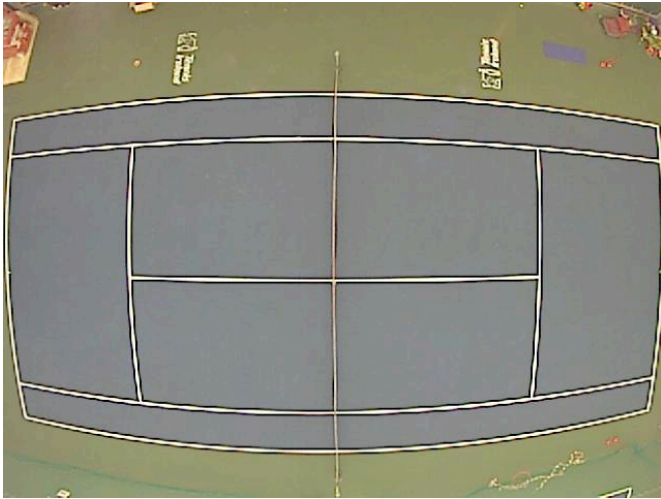
X 2 Pan+Zoom



X 7 Wide-Angle



# AV Coverage

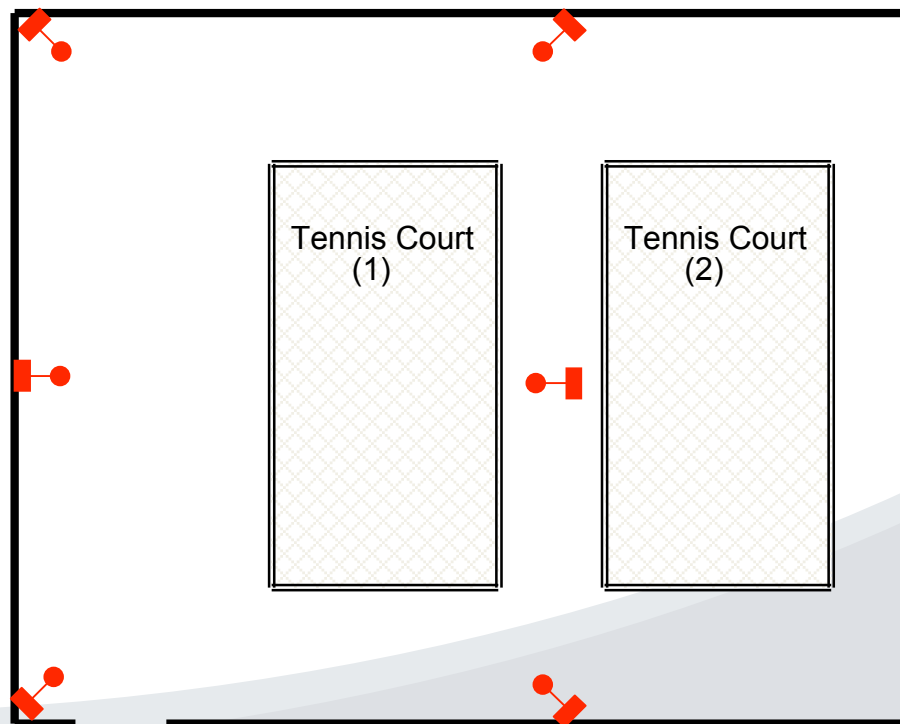




# Localisation

**Ubisense: Real-time location system, high positional accuracy**

UWB RF technology



# Sensor Pipeline

6 DOF Wearable inertial sensing platform (WISP)

Developed for Tyndall 25mm Mote

Features:

3 Accelerometers

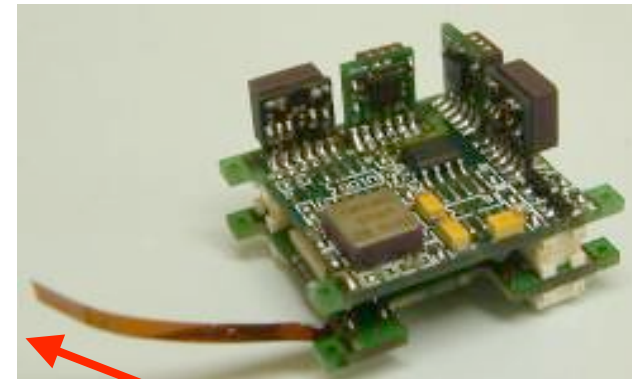
3 Gyroscopes

3 magnetometers

ADC: analog-digital converter

Microprocessor

Serial Wireless Link



25mm



10mm

# Sensor Pipeline



## Conducting polymers or “synthetic metals”

Polymers are macromolecules, and usually they are insulators but some, such as polypyrrole, conduct electricity (c.1970)

## Coat onto substrates

Foam or lycra or anything that moves, twists, bends

**Easily produced, rapid response times, comfortable to wear**

# Sensor Pipeline

**Fluid-handling platform based on polyamide lycra<sup>®</sup>**

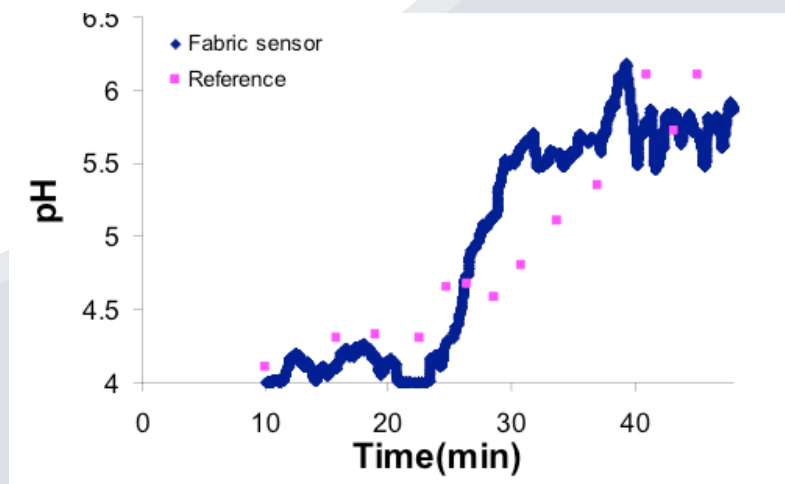
Super absorbent material (SAB) provides a passive pumping mechanism

**Used to draw sweat into its fluidic channel**

**Optical detection of pH induced colour changes in the dye**

Paired emitter-detector LED system

**Controlled by a wireless mote**



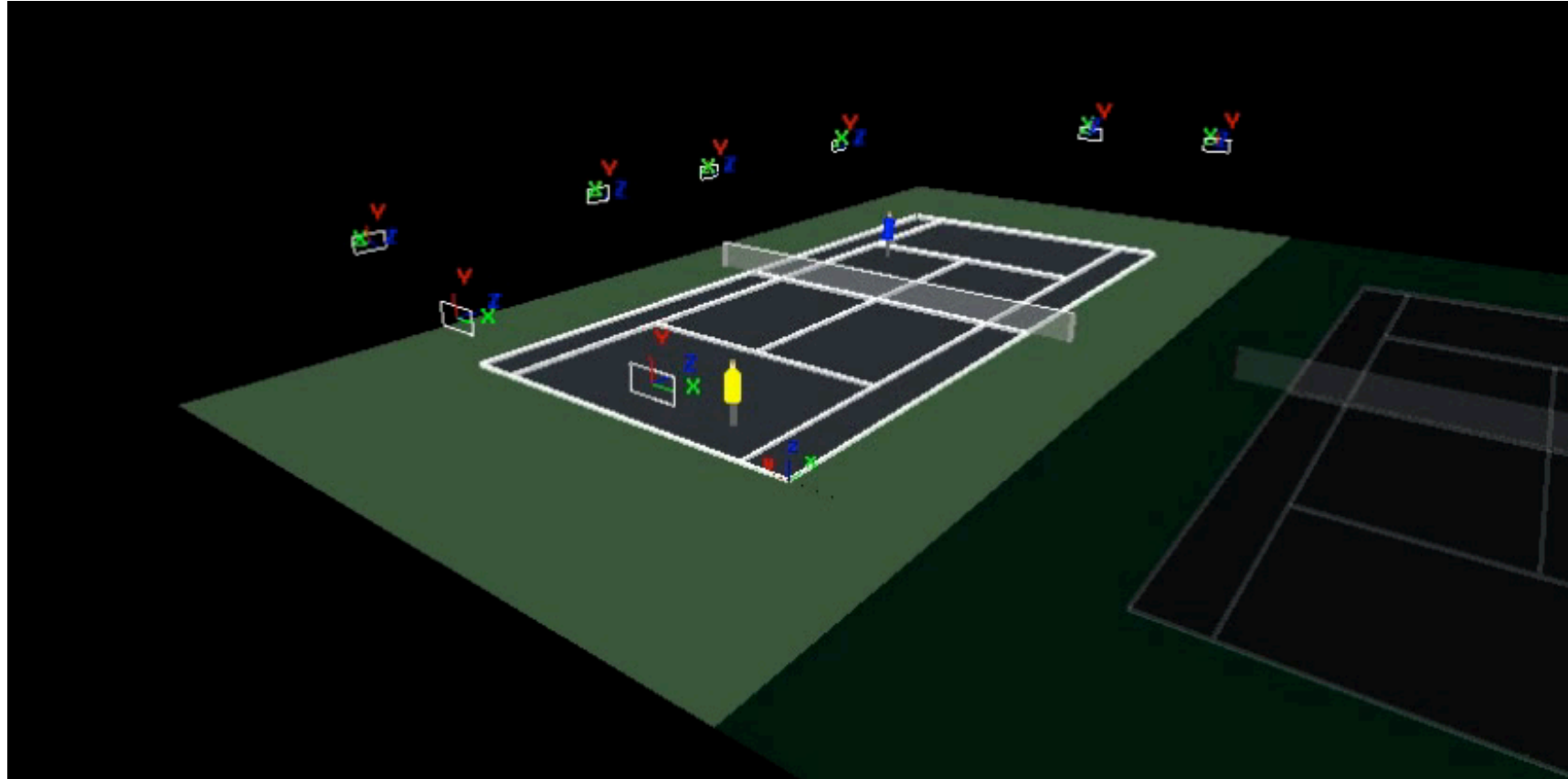
# What do people want ?

Tennis coaches, sports scientists ...

- Physiological wellness indicators in real time ... recovery between points, overall fatigue, fluid intake .. matches and training
- Strategic analysis of play for game and point winning
- Stroke analysis in training, 3D model from limb movement and WIMUs

## Roadmap ?

More integration of sensors, real time response





# Sensor App 3: Cycling

## Round Ireland cycle race

Non-stop, 1,350 miles, solo or team of 4, some hills, mostly flat

## Strategy and pacing are key to completion

When to rest, how to pace, recovery

## Human performance scientists interests

How the human body recovers, how to maximise performance over 5-day endurance  
what recovery techniques work best



# Cycle Sensors

## On Body ...

Heart rate, GSR, respiration, fluid intake, food intake, sweat analysis, blood analysis, rest and recovery techniques, massage,

## On Bike ...

WIMU gives yaw and roll, in/out saddle  
Power gauge gives power output  
Speedo gives speed & cadence  
Some direct feedback

## Environment ...

Weather, wind, temperature, RH, road surface, wet/dry, video tracking using geo-spatial measurement jeep, based on MIT



# Sensors for Cycling

## What do people want ?

Study of performance (degradation) over time, impact of different recovery techniques

## Roadmap ?

Insights into endurance performance



# Sensor App 4: Env. Monitoring

**River Lee, Cork, Ireland, a city centre tidal river**

**We can sense ...**

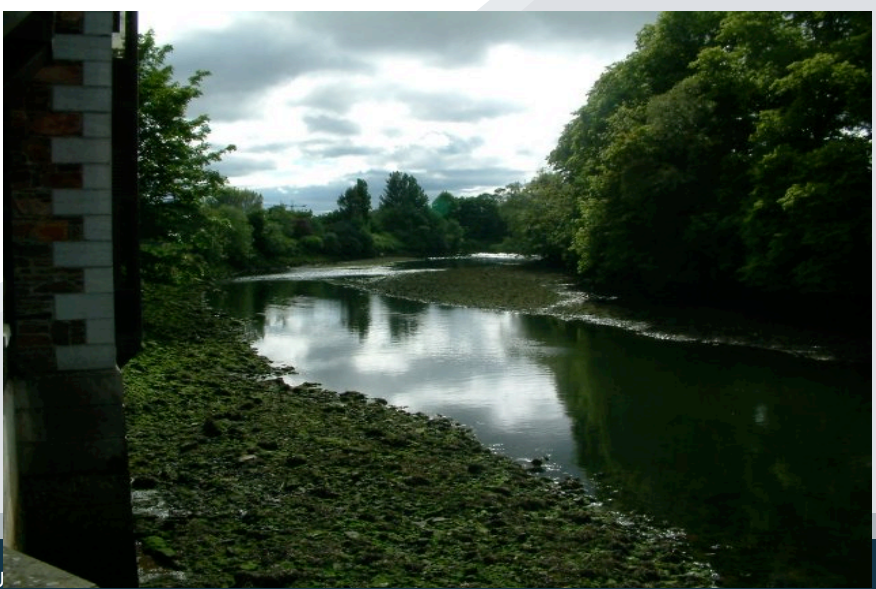
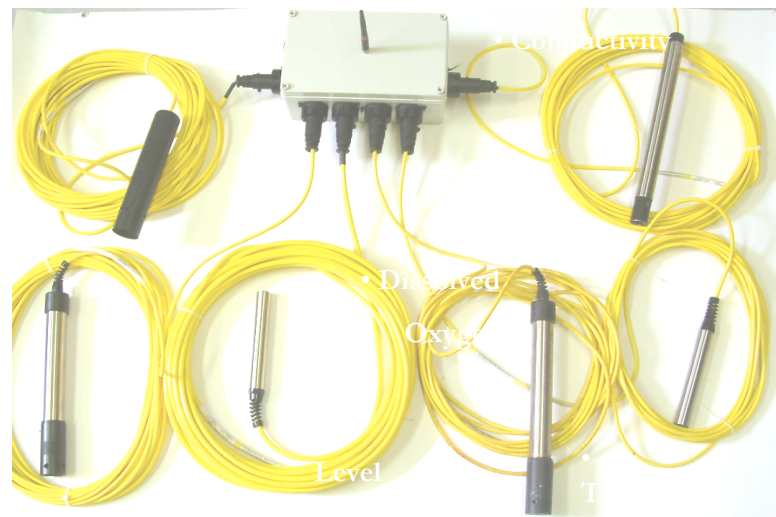
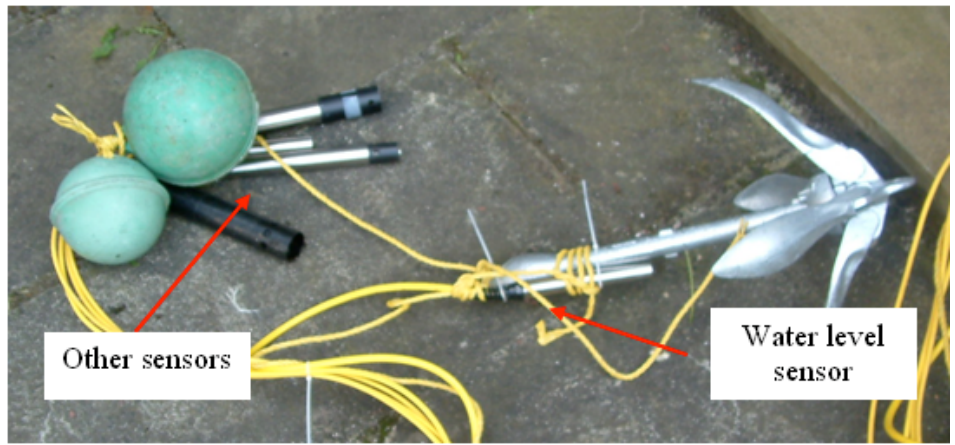
Water turbidity (NTU), water temperature (degrees Celsius), pH, water conductivity (milli-Siemens), water depth (feet), dissolved oxygen, nitrates and sulphates, air temperature, relative humidity and light (lux @ 520nm)

**Phosphates are nutrients and are bad because encourage plant growth, consume O<sub>2</sub>**

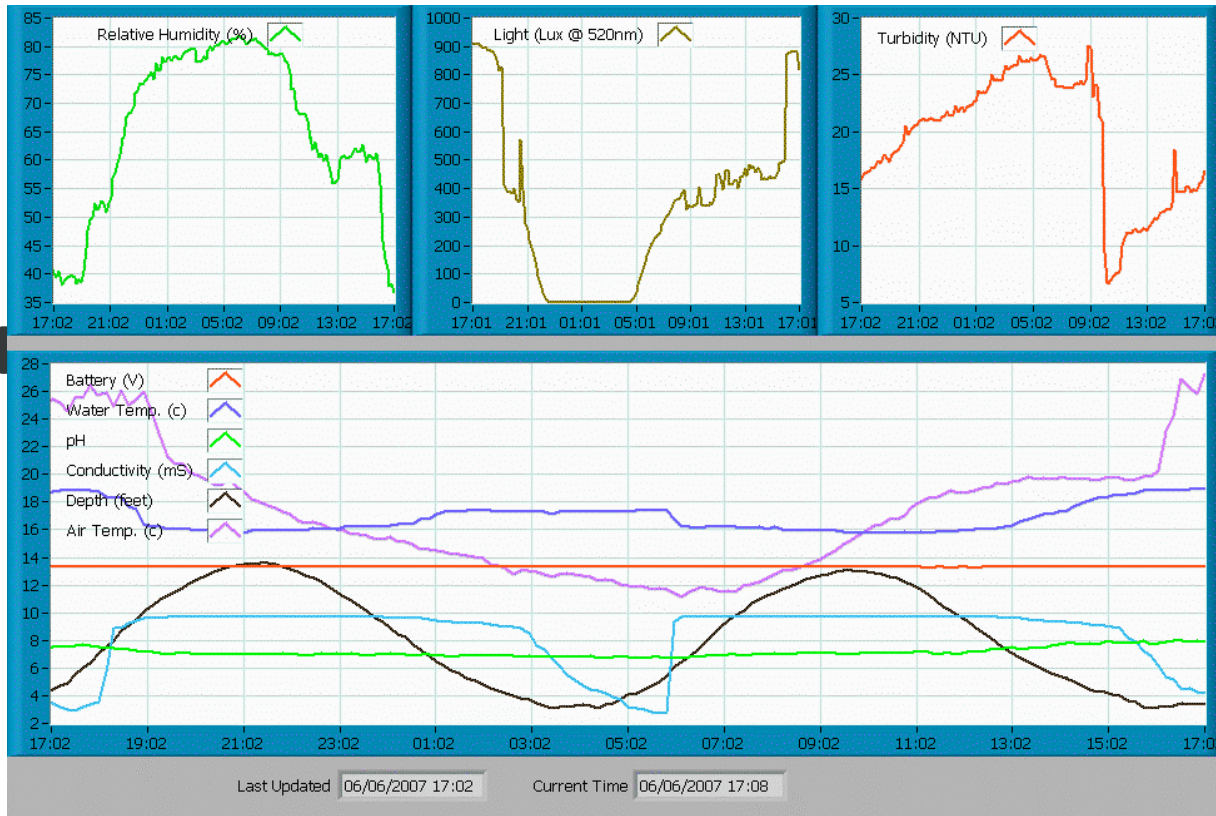
**Sensor platform is microfluidic-based, 1-year field lifetime, shoebox size, solar powered**



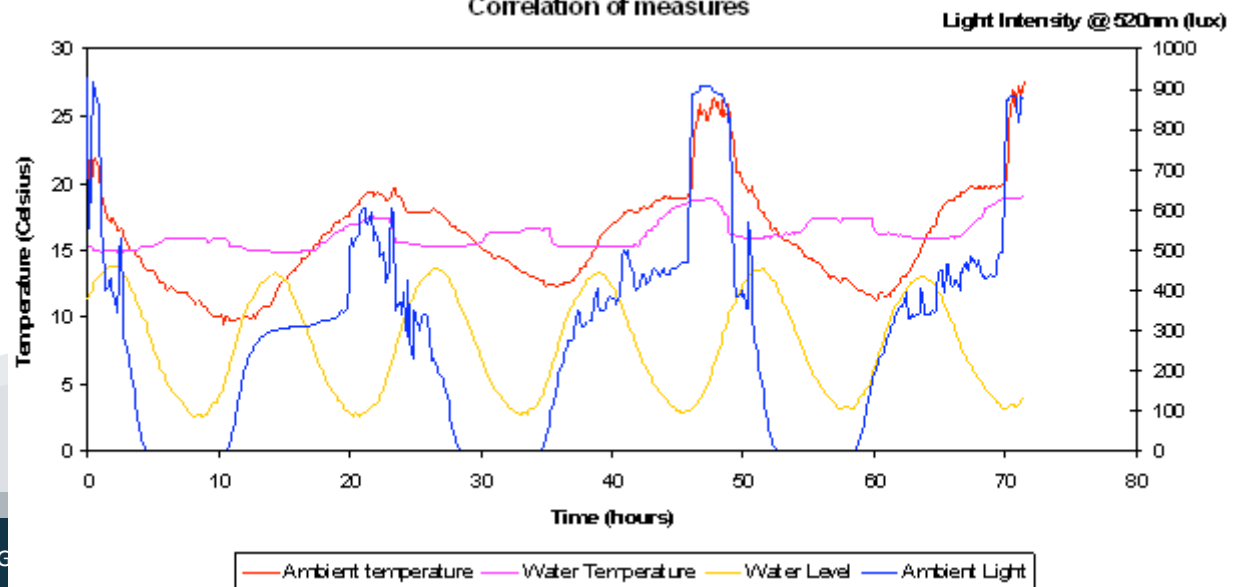
# Initial test site



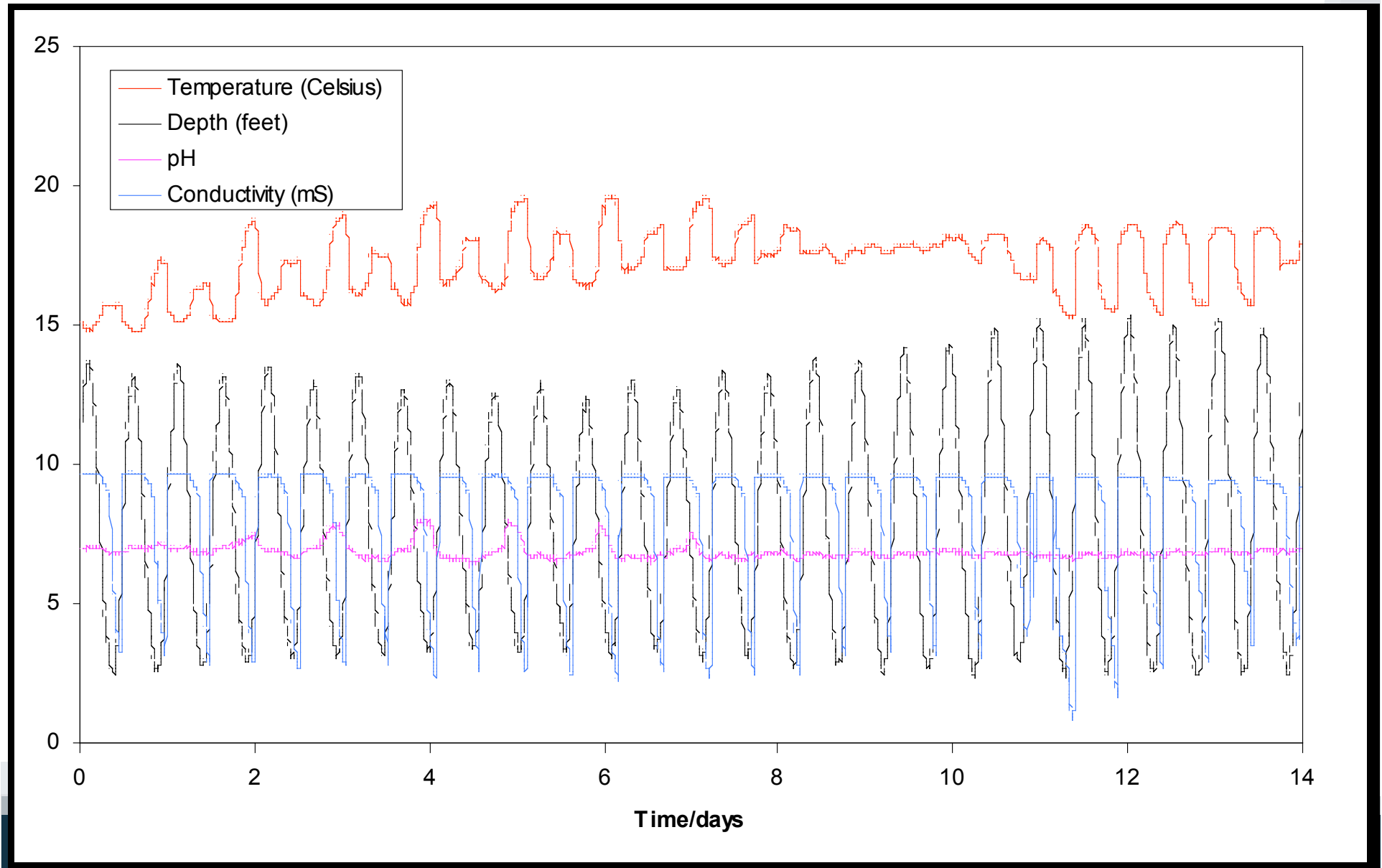




Correlation of measures



# Data Analysis, what do we do with it all?



# Other deployments ...

River Liffey (near Dublin)

River Tolka (near Botanic Gardens, Dublin)

Oberstown waste water treatment plant, Co Kildare

Black river, Newport, Co Mayo

... all measuring phosphates, nitrates, DO, as well as  
“off-the-shelf” sensors, WSN-based, ;

# Problem ... longevity

Problem with these sensors is longevity ... reagents need to be replaced

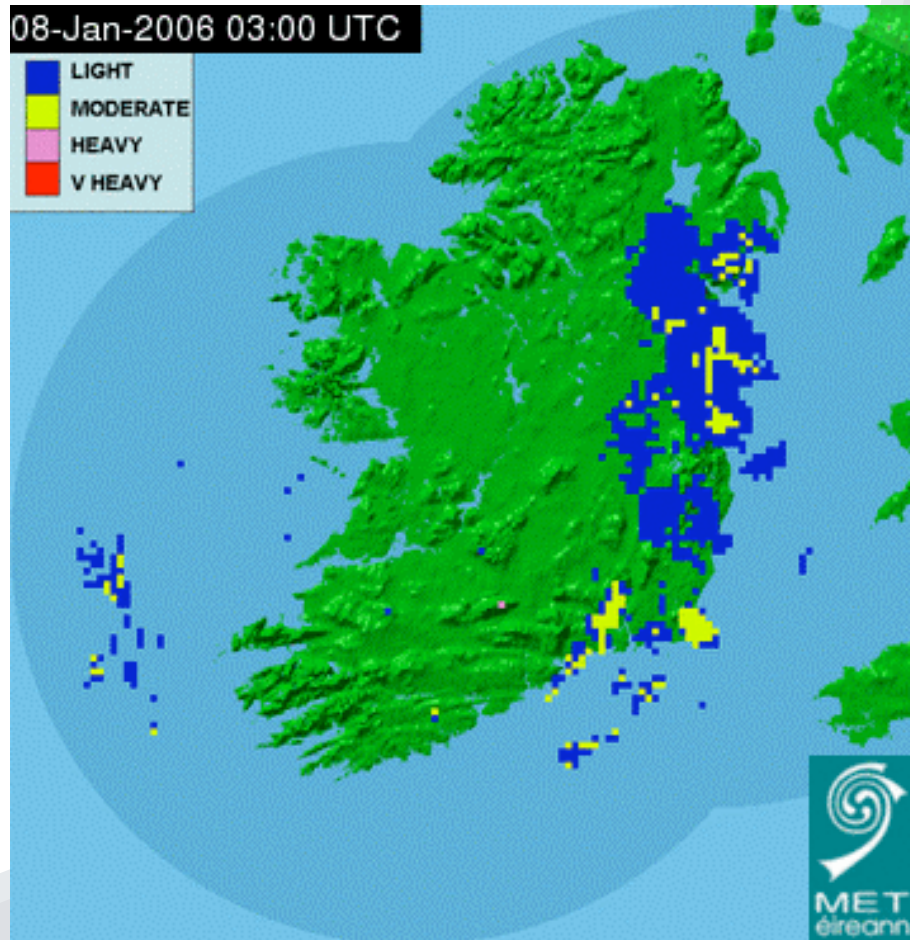
Environmental scientists want readings at times of change, not times of stability ... changing times are predictable based on rainfall

So we use another sensor ... rainfall radar

# Met Éireann Met.ie

Public website - down-sampled radar images;

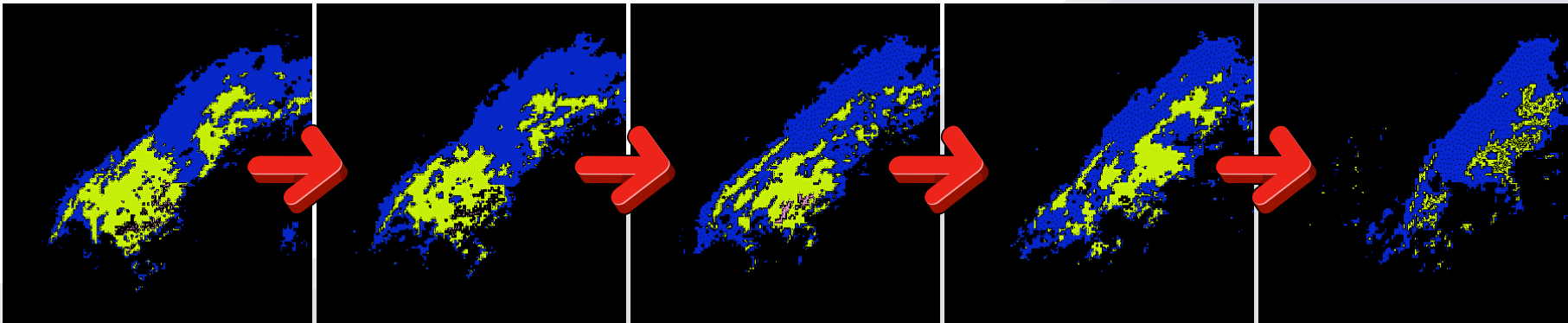
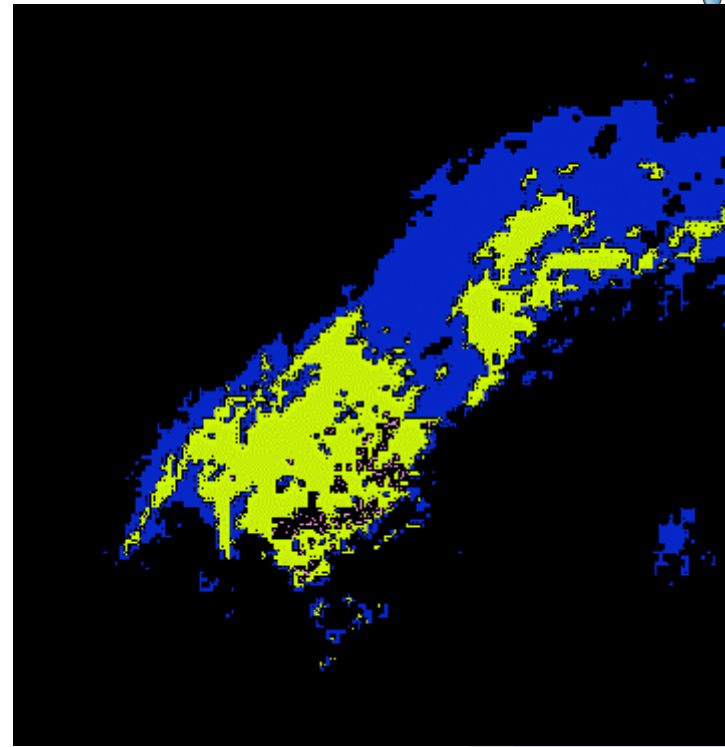
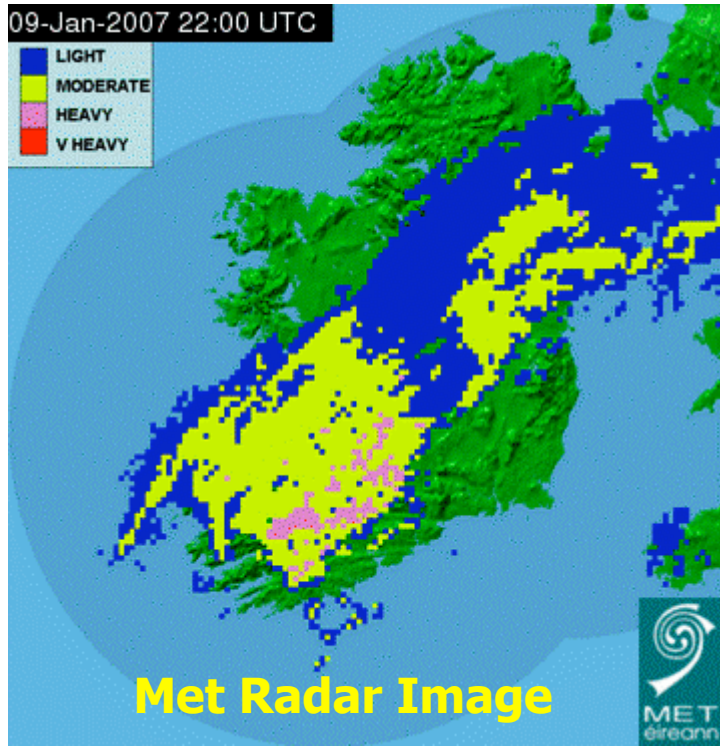
We harvest and process to track (and predict with 70% F-measure) rainfall in near-term for specific locations, the rainfall catchment areas for our river sensors;





09-Jan-2007 22:00 UTC

- LIGHT
- MODERATE
- HEAVY
- V HEAVY



# Environmental Monitoring

## What do people want ?

Long lived sensors

No biofouling !

Real time readings and prediction

Integration across sensors and sensor sources

Events !

## Roadmap ?

Smaller, cheaper sensing platforms

# CLARITY Sensor Webs

**What is common across lifelogging, tennis, cycling, and environmental monitoring ... people want semantic events**

**Events are determined by aggregating across sensors, should be done automatically**

**Where else are events automatically detected ... video**

# TRECVID goals and strategy

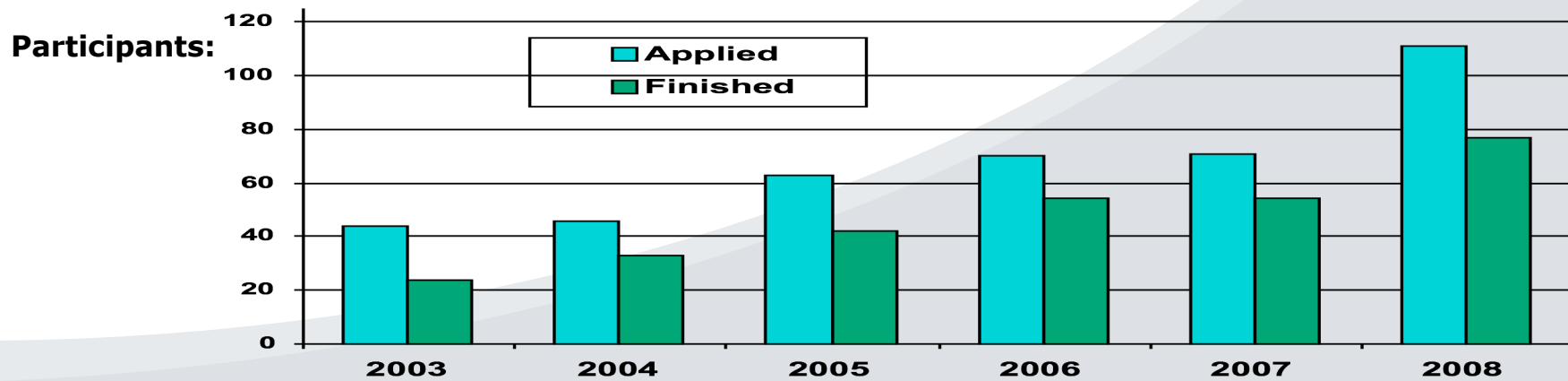
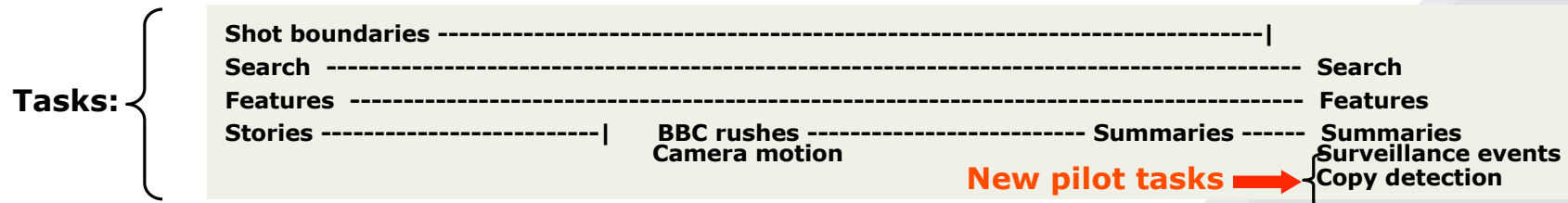
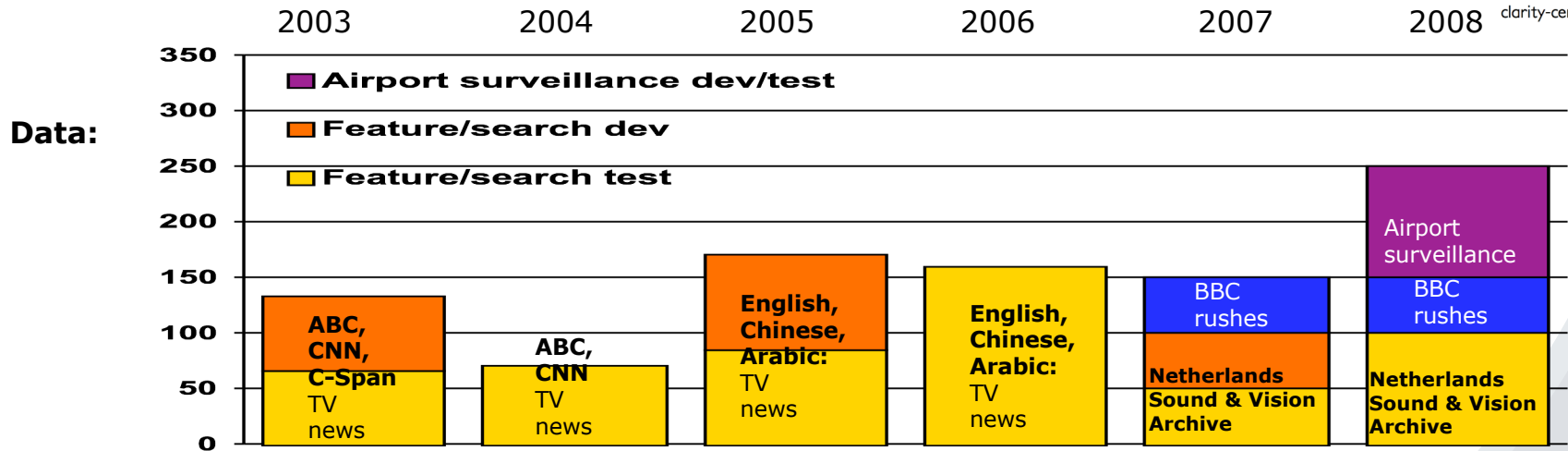
**Promote progress in content-based analysis, detection, retrieval on large amounts of digital video**

**Focus on relatively high-level functionality - automatic components:**

- Automatic search,
- High-level feature detection,
- Shot bound detection,
- Content-based copy detection,
- Event detection

**Do all this in a hugely collaborative and supportive framework, over 9 years**

# Evolution: data, tasks, participants





# TRECVID 2008: Details

## Data:

- 200 hrs - Netherlands Institute for Sound and Vision (S&V)
- 40 hrs - BBC rushes
- 100 hrs of airport surveillance data - UK Home Office

## 5 evaluated tasks

- Content-based copy detection – 2010 video queries,...
- **Semantic feature extraction - 20 features**
- Search (automatic, manually-assisted, interactive) - 48 topics
- Video summarization
- Event detection on airport surveillance video  
(5 cameras \* 2 hours \* 10 days)

# TV2008 Finishers



Athens Information Technology  
Asahikasei Co.  
AT&T Labs - Research  
Beckman Institute  
Bilkent University  
University of Bradford  
Beijing Jiaotong University  
Brno University of Technology  
Beijing University of Posts and  
Telecommunications  
Carnegie Mellon University  
Columbia University  
Computer Research Institute of Montreal  
COST292 Team (Delft Univ.)  
cs24\_kobe (Kobe Univ.)  
Dublin City University  
ETIS Laboratory  
EURECOM  
Florida International Univ.  
Fudan University  
FX Palo Alto Laboratory  
IBM T. J. Watson Research Center  
INRIA-LEAR  
INRIA-IMIA

IntuVision, Inc.  
Ipan\_uoi (University of Ioannina)  
IRIM  
ISM (The Institute of Statistical Mathematics)  
Istanbul Technical University  
IUPR-DFKI  
JOANNEUM RESEARCH  
Forschungsgesellschaft mbH  
KB Video Retrieval  
K-Space  
LIG (Laboratoire d'Informatique de Grenoble)  
Laboratoire LIRIS (LYON)  
University of Twente and CWI  
LSIS\_GLOT(CNRS LSIS)  
Marburg  
Chinese Academy of Sciences (MCG-ICT-CAS)  
Mediamill (Univ. of Amsterdam)  
MESH  
MMIS (Open Univ.)  
Microsoft Research Asia  
NHKSTRL  
National Institute of Informatics  
National University of Singapore  
National Taiwan University

# TV2008 Finishers



NTT Cyber Solutions Laboratories  
Orange Labs - France Telecom Group  
Osaka University  
Oxford Univ.  
PKU-ICST (Peking Univ.)  
PicSom (Helsinki University of Technology)  
Queen Mary University of London  
Queensland University of Technology  
REGIM  
Shanghai Jiao Tong University (SJTU)  
SP-UC3M (Universidad Carlos III de Madrid)  
The Hong Kong Polytechnic University  
Tsinghua University - Intel China Research Center  
Tsinghua University  
TNO-ICT  
Toshiba Corporation  
Tokyo Institute of Technology  
University of Alabama  
University of Electro-Communications  
University of Glasgow  
University of Karlsruhe (TH)  
University of Ottawa - SITE  
University of Sheffield

University of Southern California  
Universidad Rey Juan Carlos  
Universidad Autonoma de Madrid  
Universite Pierre et Marie Curie - LIP6  
VIREO (City University of Hong Kong)  
vision@ucf (University of Central Florida)  
VITALAS (CERTH-ITI (GR), CWI(NL),  
U.Sunderland (UK))  
XJTU (Xi'an Jiaotong University)

# Semantic Feature Detection

20 LSCOM features evaluated

1 Classroom  
2 Bridge  
3 Emergency\_Vehicle  
4 Dog  
5 Kitchen  
6 Airplane\_flying  
7 Two people  
8 Bus  
9 Driver  
10 Cityscape

11 Harbor  
12 Telephone  
13 Street  
14 Demonstration\_Or\_Protest  
15 Hand  
16 Mountain  
17 Nighttime  
18 Boat\_ship  
19 Flower  
20 Singing

# Task Context

Common shot bounds, multiple 'run' submissions allowed

Manually annotated training data available - double or triple annotation (collaborative)

Much supplemental material shared among participants

Time-bound process, c.5 weeks to complete task

Assessment based on pooling submissions and manual assessment of feature presence

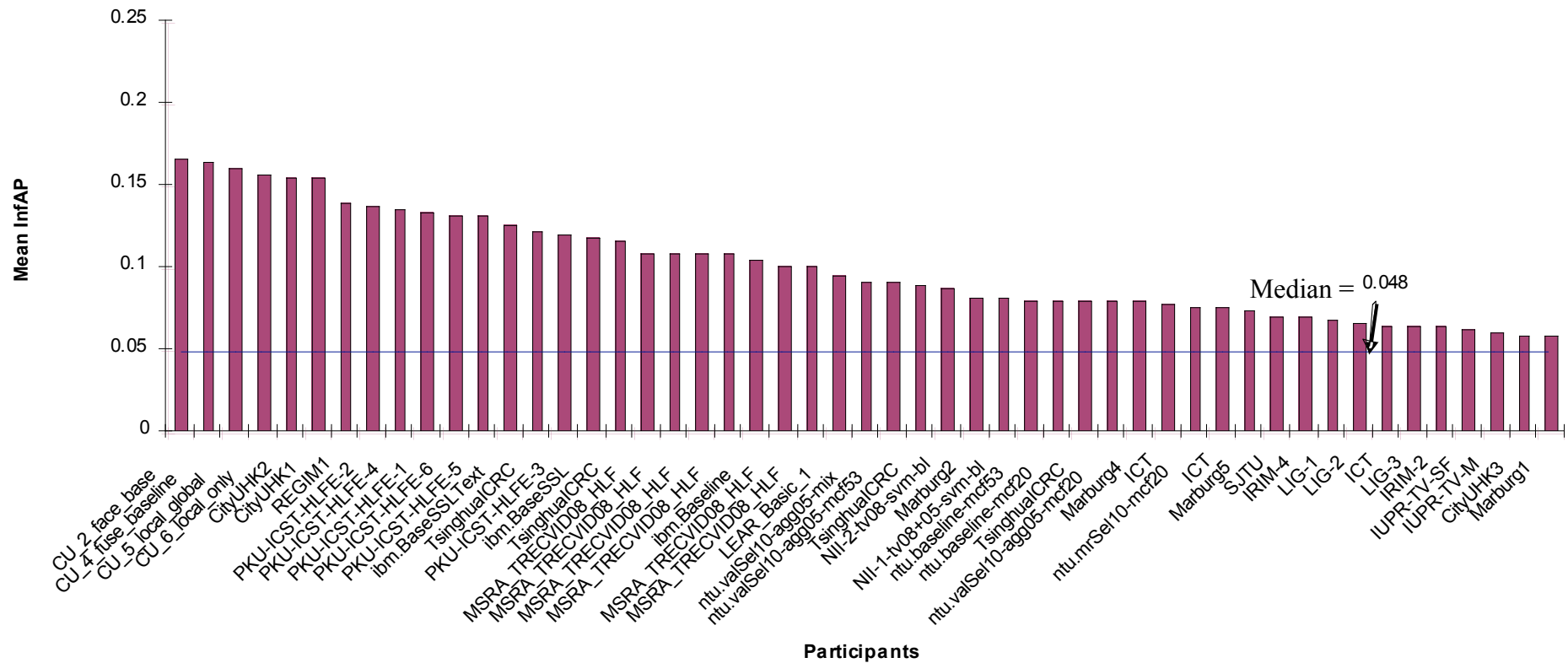
Significance testing validates results

# General observations

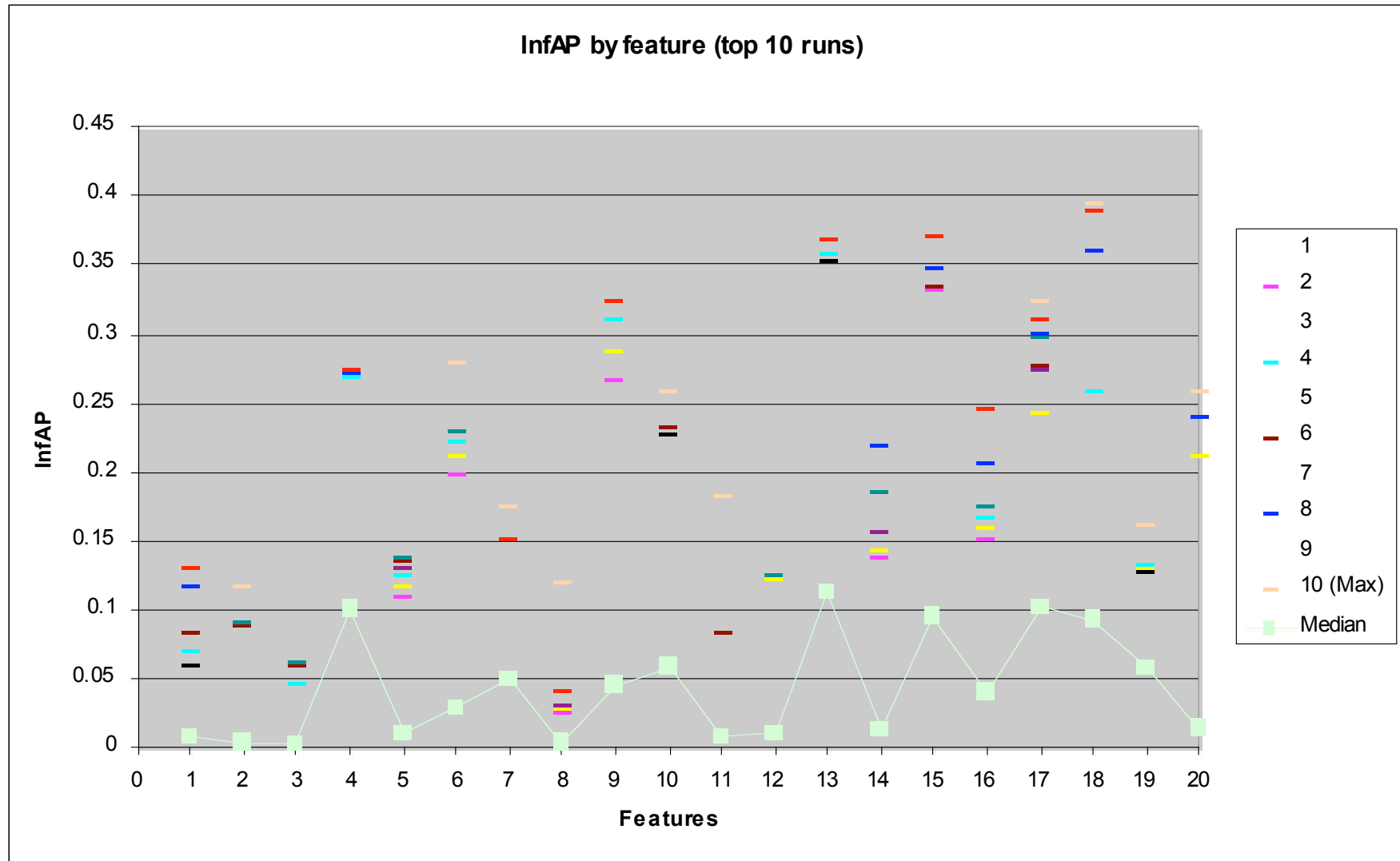
- Very popular task;
- Hardly any feature-specific approaches;
- Large variety in classifier architectures and choices of feature representations;
- Usually a single, cpu, but some medium and larger clusters;
- No. classifiers used for fusion ranges 1 .. >1160
- Times vary between 10m and 150h per feature;
- 30% of runs do some form of temporal analysis;
- 50% of runs use salient/SIFT points;
- These are features PER SHOT, or per KF - not per scene !

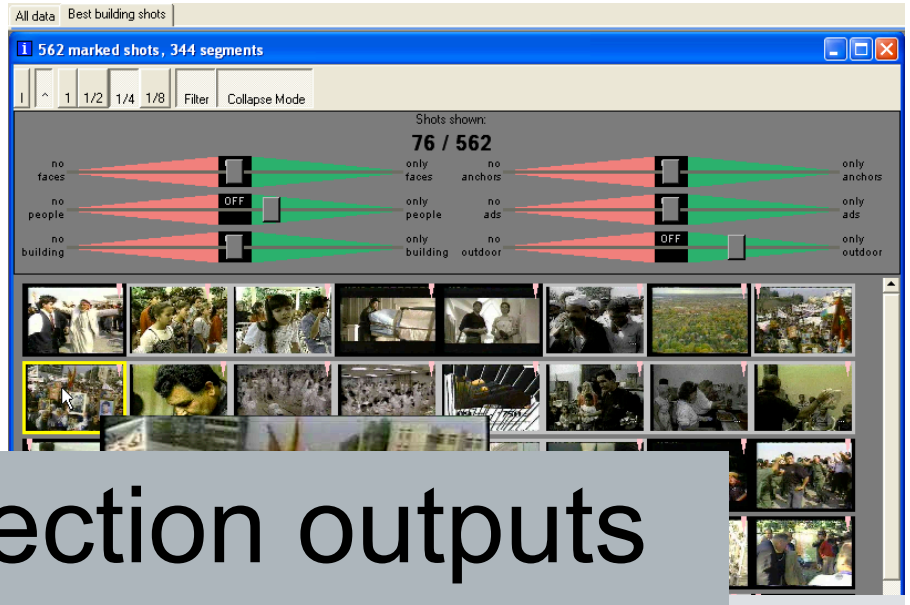
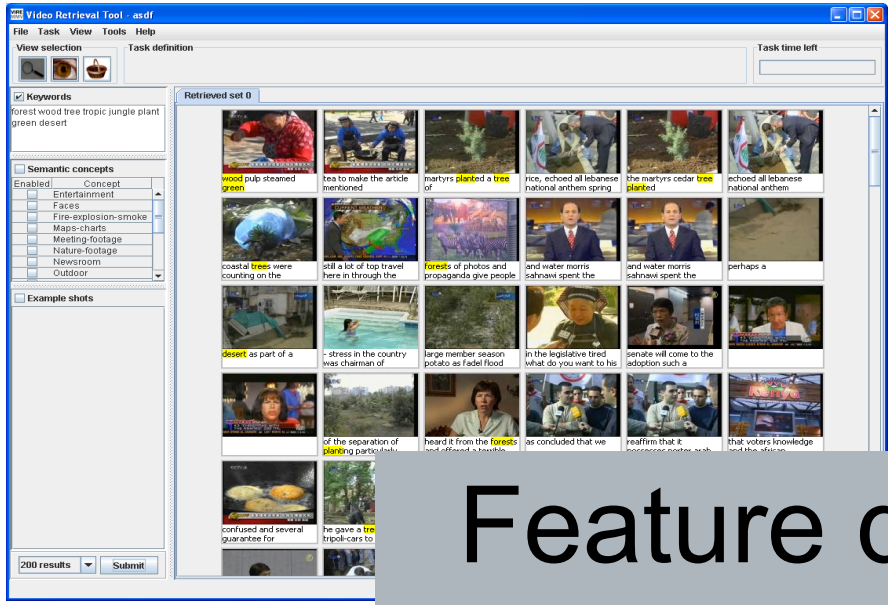


Category A results (Top 50)

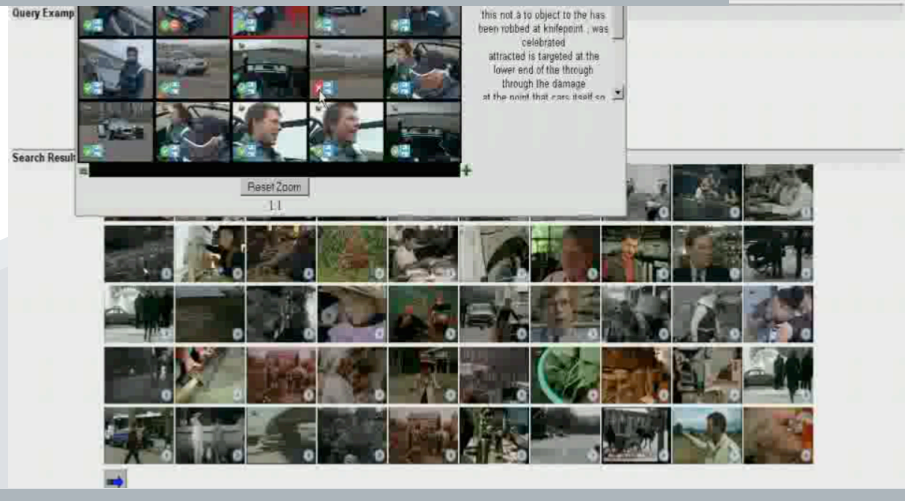
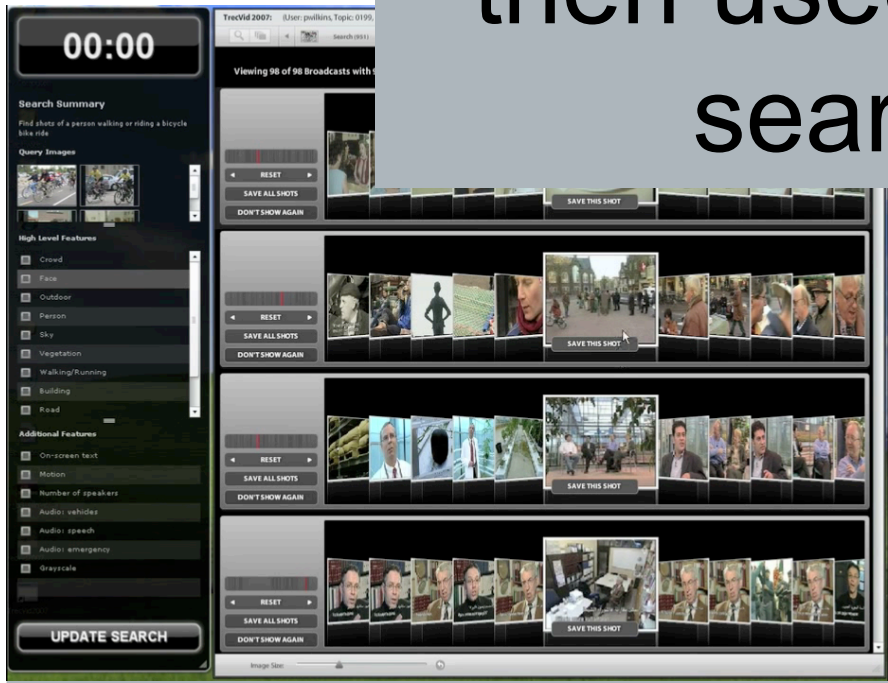


- 1 Classroom    2 Bridge    3 Emergency\_Vehicle    4 Dog    5 Kitchen    6 Airplane\_flying    7 Two people    8 Bus
- 9 Driver    10 Cityscape    11 Harbor    12 Telephone    13 Street    14 Demonstration\_Or\_Protest    15 Hand
- 16 Mountain    17 Nighttime    18 Boat\_ship    19 Flower    20 Singing





# Feature detection outputs then used in participants' search systems



# TRECVID Summary

On small, closed video libraries, content based video search works well; with metadata and UGC it would be even better ...

We're still only doing keyframe/image and tasking shot retrieval and we're purposely not using metadata or tags or UGC;

**Feature detection accuracy, scale-up to more features, relationships between features, move away from independence to ontology-based ... need to progress this;**

Combining concepts, keyframe match, text and objects for searching in a natural and usable way;

## What do people want ?

Better accuracy, more concepts, concept dependencies, diversity in video sources

# TRECVID Summary

On small, closed video libraries, content based video search works well; with metadata and UGC it would be even better ...

We're still only doing keyframe/image and tasking shot retrieval and we're purposely not using metadata or tags or UGC;

**Feature detection accuracy, scale-up to more features, relationships between features, move away from independence to ontology-based ... need to progress this;**

Combining concepts, keyframe match, text and objects for searching in a natural and usable way;

**What do people want ?**

Better accuracy,  
sources

**Sound Familiar ?**

encies, diversity in video

# Earlier I introduced ...

## The 'Old' Web

## The Social Web

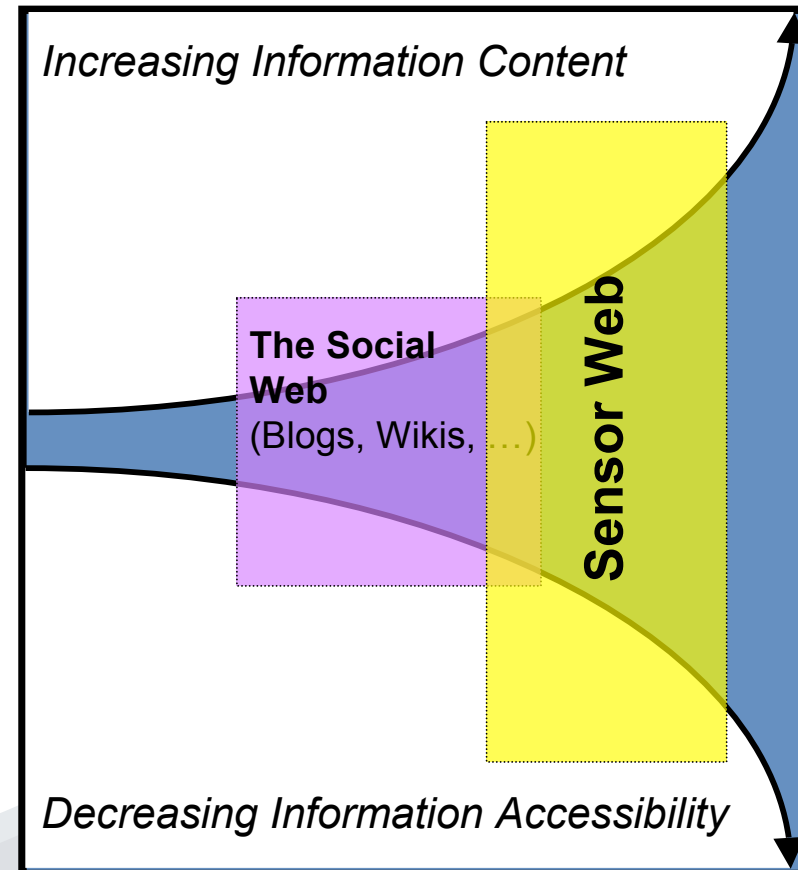
Dynamic user generated content (UGC) - conventional search tools unable to cope

## The Semantic Web

In here somewhere ...

## The Sensor Web

Increasing availability of cheap, robust, deployable sensors as ubiquitous information sources  
Key is that they are networked, mostly wireless, hence global, and integrated - WSNs  
Dynamic and reactive but noisy, and unstructured data-streams





# The Sensor Web

Mooted because sensors for local control surpassed by networked sensors, giving networked, online hotspots;

Vision of a global, wholly integrated, WSN web ... not here anytime soon;

We saw examples of local sensor networks with SN characteristics ... noise, errors, calibration, etc.

We also saw examples of what users want, which drives sensor development

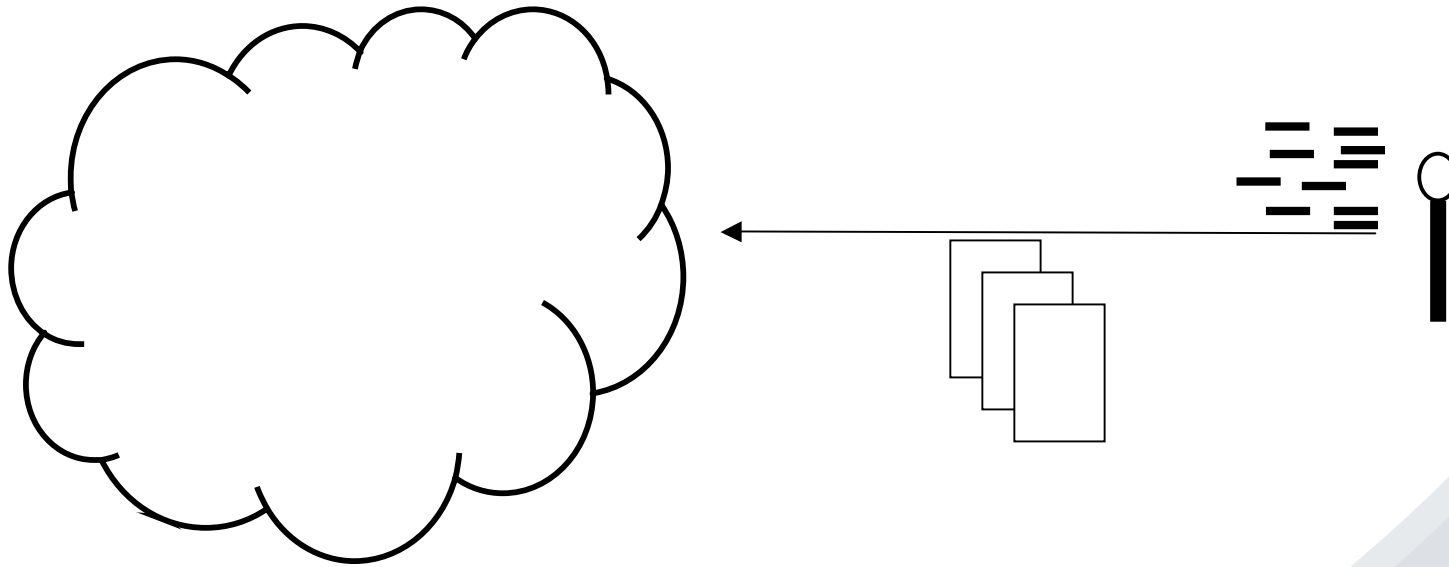
# Sensor Web/Net Characteristics

## Common characteristics ...

- Niche domains
- Known, pre-defined events to be detected
- No single best detection solution, no rules
- Detection accuracy not always great, but workable
- Limited uses of concept ontology

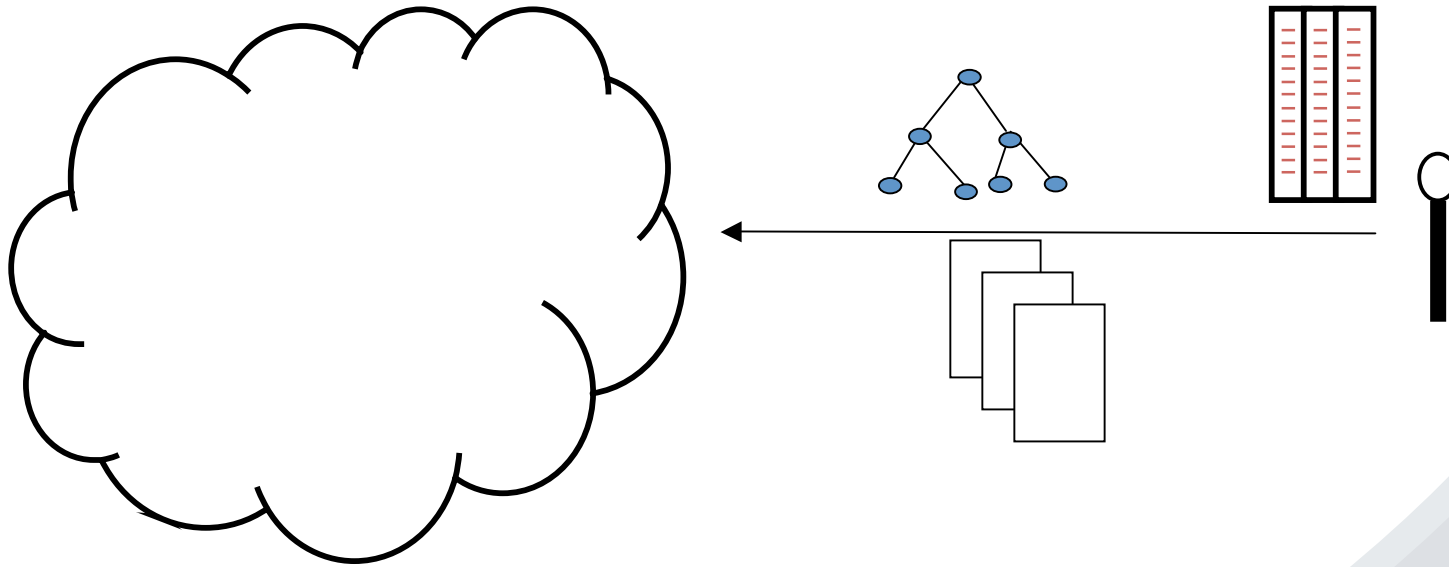
**These are precisely the same characteristics as concept detection in video - a surprising parallel**

**And where does the sensor web fit in ?**



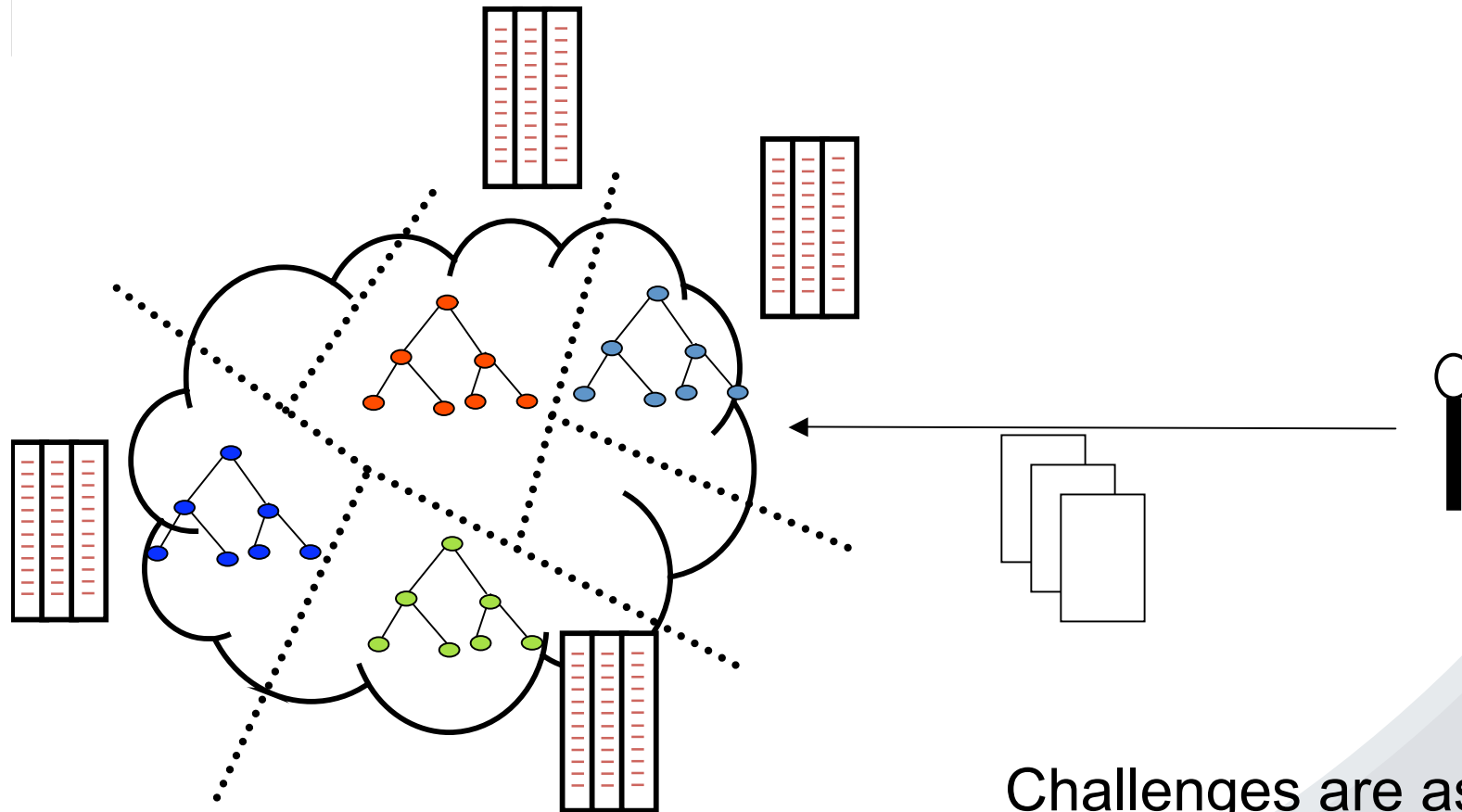
The 'old' web !

Challenges: scale,  
polysemy, NLP



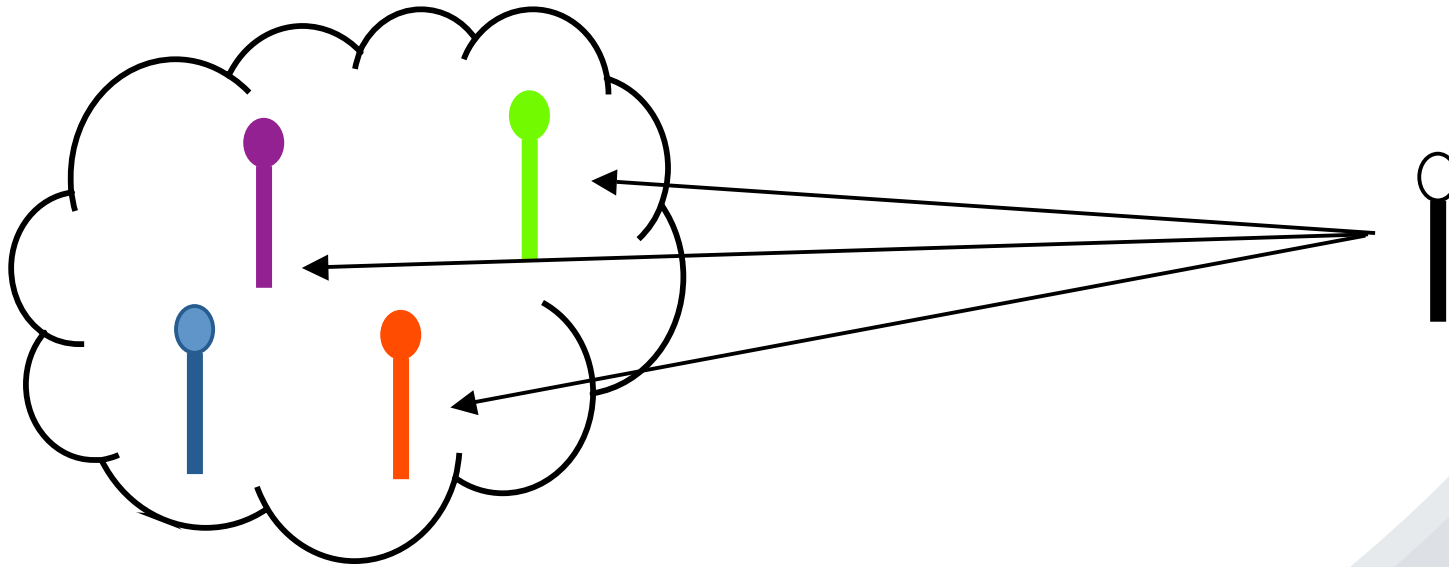
The semantic web

Initially ... one SW to  
bind them all ... now ...



The semantic web

Challenges are as before  
... + ...  
ontology creation, etc.  
content parsing  
scale and processing  
inference and reasoning

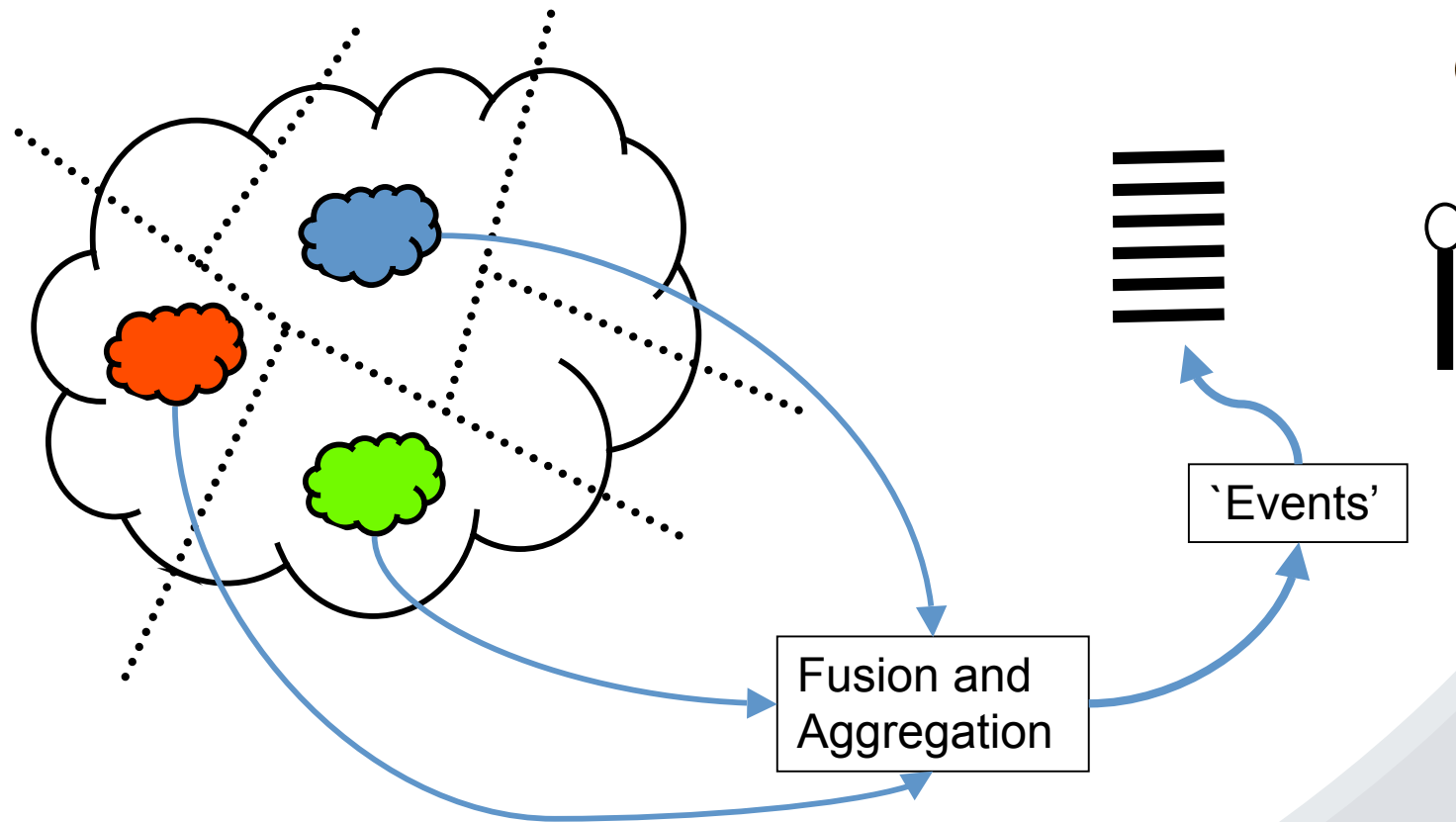


The social web

Challenges are ...

- people-centric
- information diversity  
(source, media, sentiment)
- archive to realtime info.





The sensor web(s)

Challenges are ...

- errorsome data
- fusion and aggregation
- event detection, not always deterministic and rule-based

# Thanks to ...

**Very many people in Dublin ... colleagues,  
students, collaborators**



**Science Foundation Ireland**

**... the organisers for inviting me**