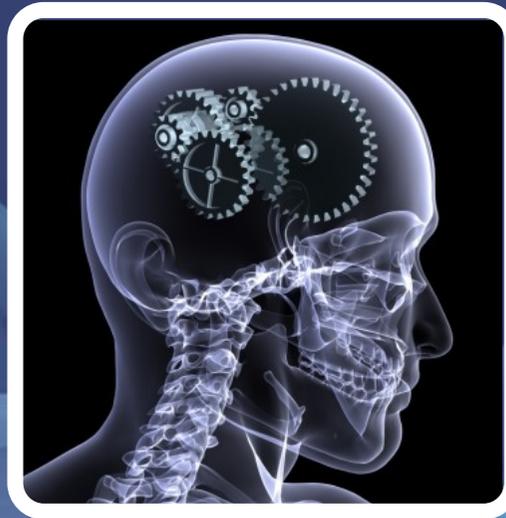# Information Retrieval Challenges of Maintaining a Context-aware Human Digital Memory

Dr. Cathal Gurrin
Centre for Digital Video Processing, Dublin City University, Ireland.

Centre for Digital Video Processing

- 6 Faculty, 17 PostDoc, 4 RA, 20 PhD
- CDVP Summary Research Interests:
  - Digital multimedia analysis and organisation
    - Video : TV, News, Educational, CCTV
  - Personal Media Management
    - Human Digital Memories
    - Personal Photo Management
  - Sensor Networks & SensorWeb Technologies
  - Text IR / Search Engines

Focus of this Talk

Human Digital Memories

- A Human Digital Memory (HDM) is a surrogate of your own memory, though in digital form.
- Although it may sound like Science Fiction, it is being done now, though not quite as intrusively as you would think.
- So
  –
  –
- So
  vac   ns;
- Sometimes its because we can, and we are still researching how we can use it

> "Some day we will be able to record everything we see and hear."
>
> Bill Gates, in "The Road Ahead", 1995

How to construct a HDM

- **Humans have been finding ways to substitute memory for a long time**
  - Druids, books, computers…
- **For HDMs we can…**
  - Logging our environment;
    - Traffic, pollution, **people near us**, etc…
  - Logging our **location**
    - My movement log is since November 2005
  - Logging what we see and do
    - **Visually**, aurally,
    - **from our point of view** and others' view of us
  - Recording biometrics;
    - Logging our health, fitness, activity, etc…
  - Logging our actions
    - Communications
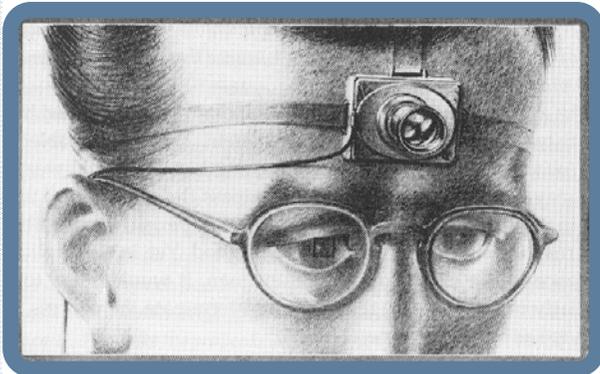   - Data interactions

?

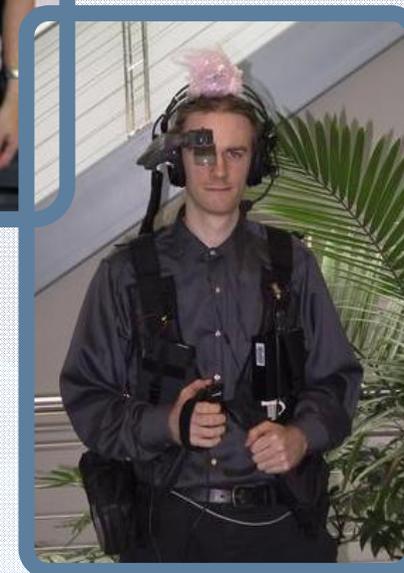Cathal Gurrin (2008)
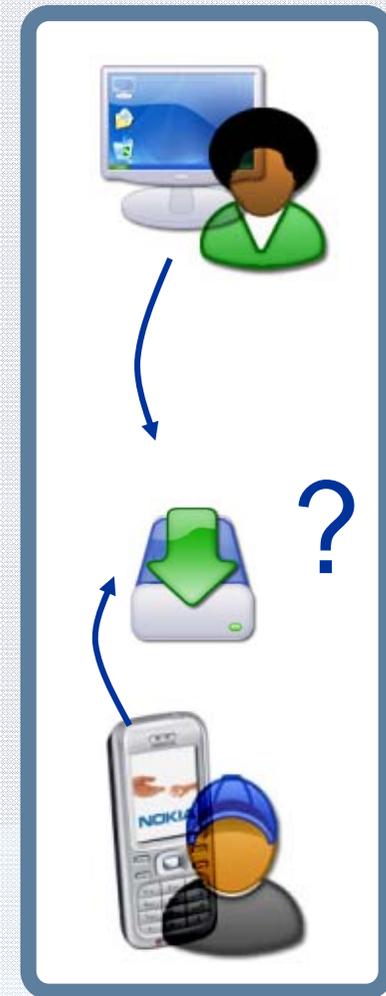
MEMEX – the first HDM scenario
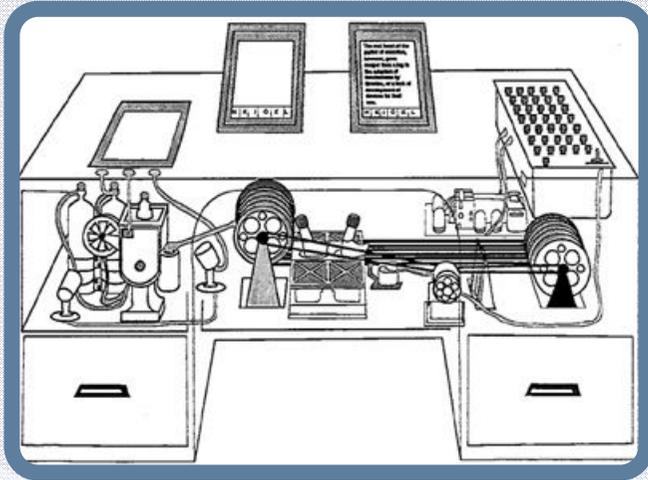
In 1945, Vanevar Bush wrote 'As We May Think'… a prophetic view of computing technology:
- Hyperlinks & WWW
  - Modelling associative memory
- Miniature wearable camera
- Lifelogging
  - Gordon Bell (Microsoft) and MyLifeBits

People have worked on this…

In this talk...

- **I will focus on the visual aspects of a Human Digital Memory**
  - Capturing what we see, and organising this data
- **Agenda**
  - Prior Research on Personal Media Archives
  - Constructing a Human Digital Memory
  - Organising a Visual Human Digital Memory
  - Challenges for Human Digital Memories
  - Human Digital Memories, My thoughts

Prior Work in Personal Media Archives

Personal Photo Collections
Personal Video Collections

Managing Personal Photos

- Most photo search relies on date/time and manual annotation/organisation
- Recently Location has been used to support search ... e.g. Flickr
- Some key points for automatic organisation
  - Photo capture is bursty in nature
    - Helps EVENT segmentation
  - Context of capture can drastically reduce the search space
    - Date/Time/Location/People/biometric data
  - Content Analysis can also help (feature detectors)
    - Baby/dog/face/building

MY FRIENDS' FACES

Search and browse people's faces in your photos and share with friends in a fun and easy way

On the table

Dublin, Ireland
5 June 2005

Owner: **Hyowon** U
834 Photos in **106** Events

Search for people, location, or event in annotation

[ ] GO ●

ADVANCED »

● PEOPLE LIST

86 people annotated. To search by person, select & click GO button (Hold Ctrl for multiple selection).

▼ NAME	PHOTO	EVENT
Fabrice	23	3
Paul	234	43
Georgina	76	20
Cathal B	1,430	95
Mary	15	3
Hyowon	9	2
Bart	53	6
Neil B	1,316	103
Alan B	635	25
Sandra	128	14

▼ MORE

● GO

RECENT EVENTS ●

FAVE

INDOOR 🅘 CITYSCAPE ◐

PUBLIC ●

On Cathal's birthday party on Friday afternoon in The Arc, Liffey Valley. People were happy and day was bright, restaurant was pleasant. We liked the day very much! The meal was also very nice.

The Arc: http://www.thearc.ie 🔖
Dublin tourist guide: http://www.dublin-tour.ie 🔖

COMMENTS

Joyce U
Nice photo! I hope you guys enjoyed the event.
(7 Jun 2005)

Christopher U
Very nice shot with nice lighting! What restaurant is this? Let me know if you visit this place again..
(22 Jul 2005)

Cathal B
Ah, thanks for uploading this photo! (23 Jul 2005)

↳ **Hyowon** U
No problem, I had it for a long time but hadn't uploaded. (24 Jul 2005)

Aoife U
I like the colour - well done! (1 Aug 2005)

Dave U
Ireland must be a good place for celebrating a

PEOPLE IN THIS PHOTO

Paul
Georgina
Cathal B
Mary

● NAVIGATE EVENT

Cathal's birthday Party (27 Photos)

● PEOPLE LINK

See these people in other photos in:
This Event | **All My Photos** | **All My Buddies**

Paul (234 Photos)

Georgina (76 Photos)

Cathal B (1,430 Photos)

MORE ▼

● EVENT LINK

See similar Events in:
My Photos | **All My Buddies** | **All Users**

MORE ▲

Simon U
EVENT **Dublin**, Aug 06 (24 Photos)

EVENT **Dublin**, Feb 06 (72 Photos)

Joe U
EVENT **Dublin**, Apr 06 (59 Photos)

What we have learned?

- Event segmentation is important to organise the content
  - For presentation, search and feature inference.
- Context is key for accessing content based on a person's memory of the event
  - A person's query will be incomplete/incorrect
  - When an event occurred is not easily recalled
    - Where the event took place is!
- Content analysis tools will inevitably fail some/much of the time, so need to be able to manually fix.
  - When they work, they can be very useful if the correct suite of tools are chosen
  - What are the correct tools for HDMs?

Personal VIDEO Archives

Integration of Content Analysis, specifically visual feature extraction tools

Supporting Search in Video

- Video data organised into video shots and scenes for easier organisation
- Text Search
  - ASR, CC, OCR, manual annotations (e.g. EPG or community annotations) to generate surrogates of the shots and scenes
- Visual Content Search
  - Automatically extract features from each keyframe/shot/...
    - Image/Video low-level features
    - Derived Attributes
      - People, Location, Objects, Events
      - Likely trained using an SVM over the low-level features
  - CMU estimate that about 5,000 concepts detected with minimum accuracy of 10% is needed to provide web search quality results for domain limited video

QUERY PANEL

Enter term(s) and click SEARCH button. Added shots will be used together for searching.
[clear text] [clear all]

rocket launch   SEARCH

☐ GLOBAL COLOUR   ☐ LOCAL COLOUR
☑ EDGE   ☑ MOTION
☐ TEXTURE   ☐ FACE FILTER ☹

had failed repeatedly and publicly and there're tests...

REMOVE ✗

powerful rockets that in nineteen sixty one launched during...

REMOVE ✗

are shuttle launch can be built in wrestling but...

REMOVE ✗

SEARCH RESULT

Search found **250** matching shots. Following is the **ranked list** of the search result. Click an image to play the segment, click **SAVE ▶** button to add to your answers, or click **◀ QUERY** button to improve subsequent query.

RESULT PAGE : 1 2 3 4 5

1: ABC News (30 Oct 1998)   > MORE MATCHES IN THIS BROADCAST

◀ QUERY | SAVE ▶   ◀ QUERY | SAVE ▶   ◀ QUERY | SAVE ▶   ◀ QUERY | SAVE ▶   ◀ QUERY | SAVE ▶

that manages about we've got big he'd be in on board discovery yesterday to **launch** would have looked like this when is in the middle of this picture are shuttle **launch** can be built in wrestling but it was far gentler than lives at least one we've that the astronauts back in space again that was quite different than a ride i got before of course back a long time ago

2: CNN News (26 Oct 1998)   > MORE MATCHES IN THIS BROADCAST

◀ QUERY | SAVE ▶   ◀ QUERY | SAVE ▶   ◀ QUERY | SAVE ▶   ◀ QUERY | SAVE ▶   ◀ QUERY | SAVE ▶

those were the coldest years of the cold war the soviet military arsenal included powerful rockets that in nineteen sixty one launched during the government the first person in space this all u. s. **launch** rockets

3: CNN News (26 Oct 1998)   > MORE MATCHES IN THIS BROADCAST

◀ QUERY | SAVE ▶   ◀ QUERY | SAVE ▶   ◀ QUERY | SAVE ▶   ◀ QUERY | SAVE ▶   ◀ QUERY | SAVE ▶

the government the first person in space this all u. s. **launch** rockets had failed repeatedly and publicly and there're tests pressure was building dev... american investments these are the past there have been

4: CNN News (26 Oct 1998)   > MORE MATCHES IN THIS BROADCAST

■ SAVED SHOTS

You saved **2** shots, shown below - these are your **answers** to the current task. Add more, delete, or use them for subsequent query.

◀ QUERY

DELETE ✗

◀ QUERY

DELETE ✗

▶ ○ ■ ◀ ▶ 🔊

What have we learned?

- **Shot Boundary Detection and scene detection is important to organise the video content into meaningful units**
  - For presentation, search and feature inference.
- **Content analysis tools can help a lot**
  - Either a lot of tools (,000s) or carefully chosen domain dependent tools (e.g for sports summarisation)
  - Are the same HDM tools suitable for HDMs?
- **Textual surrogates (not visual content) are still are key access and organisation methodologies**
- **Keyframe size and layout help to guide a user to the most important parts of the content**

Constructing a Human Digital Memory

Movement
People we meet
Our Interactions
What one sees

Logging Movement
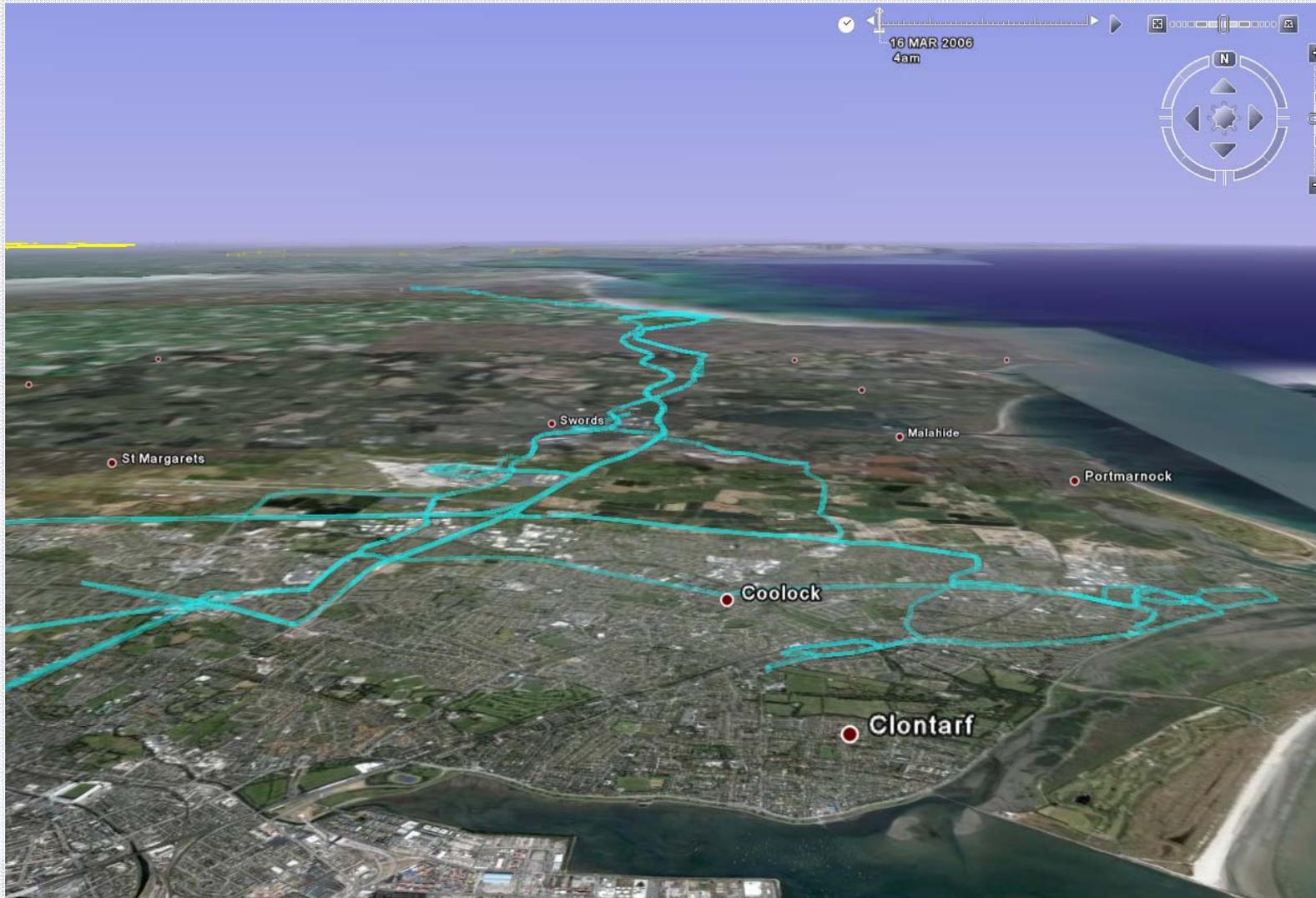
Logging People Interactions

- ## Social Network Generation
  - Based on real-world interactions using Bluetooth on mobile devices

- ## This allows us to log who is near to us at any one time

- ## Can help to organise our HDM by automatically annotating events with people likely to be there

Logging our Actions

- **Biosensors can log our actions and even log aspects of our emotion**
  - Require wearing sensors and not likely to be comfortable to wear
    - E.g. heart rate, excitement, respiration,…
  - Sensecam gathers environmental temperature
- **Logging a person's communications and data actions**
  - Phone calls
  - Text messages
  - Emails
  - Web pages visited
  - Content of documents worked on

Logging what we see...

- Capture digitally what a person sees and experiences over a period of time
- Envisaged by:
  - Vannevar Bush in 'as we may think' as a Walnut sized camera on one's forehead
  - Bill Gates in "The Road Ahead"
  - Many others and in many books and movies
- This has become a reality, lead by projects such as MyLifeBits, Sensecams, and Mr Lee, the lifelogging cat.

SenseCam: Logging what we see

- ## SenseCam is a Microsoft Research Prototype
- ## Multi-sensor device:
  - colour camera (vga with fisheye lens)
  - Sensors
    - 3 accelerometers
    - light meter
    - passive infrared sensor
- ## 1GB flash memory storage of over a week
- ## Smart image capture ~3 images/min
- ## CDVP have seven Sensecams
  - We see them as general HDM creation tools
 - Other research is more domain specific (e.g. health)

SenseCam Photos

So, what does a day look like?

Analysing the Visual HDM

- Wearing a Sensecam will produce over a million photos in one year
  - Assuming about 16 hours per day
- These photos:
  - Have huge levels of redundancy
  - Vary in quality from unusable to photo-album
  - Captured scenes differ from conventional photos
    - Non-standard photo capture environments, like inside a car, or in my office.
  - Don't typically have a salient object
    - Sometimes people are captured 'headless'
  - Capture the hands of the wearer

Analysis: Redundancy

- Huge redundancy in visual HDM data.
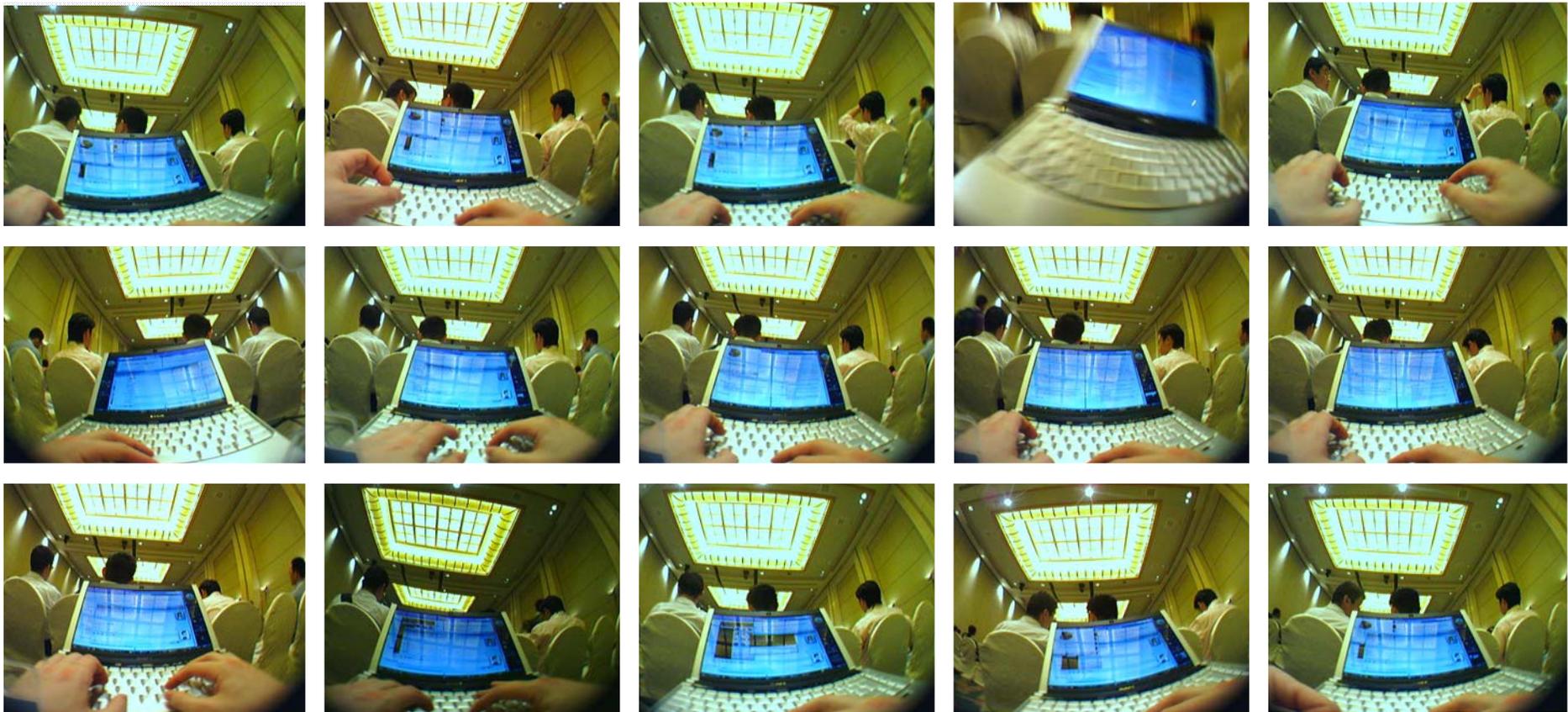
Analysis: Quality
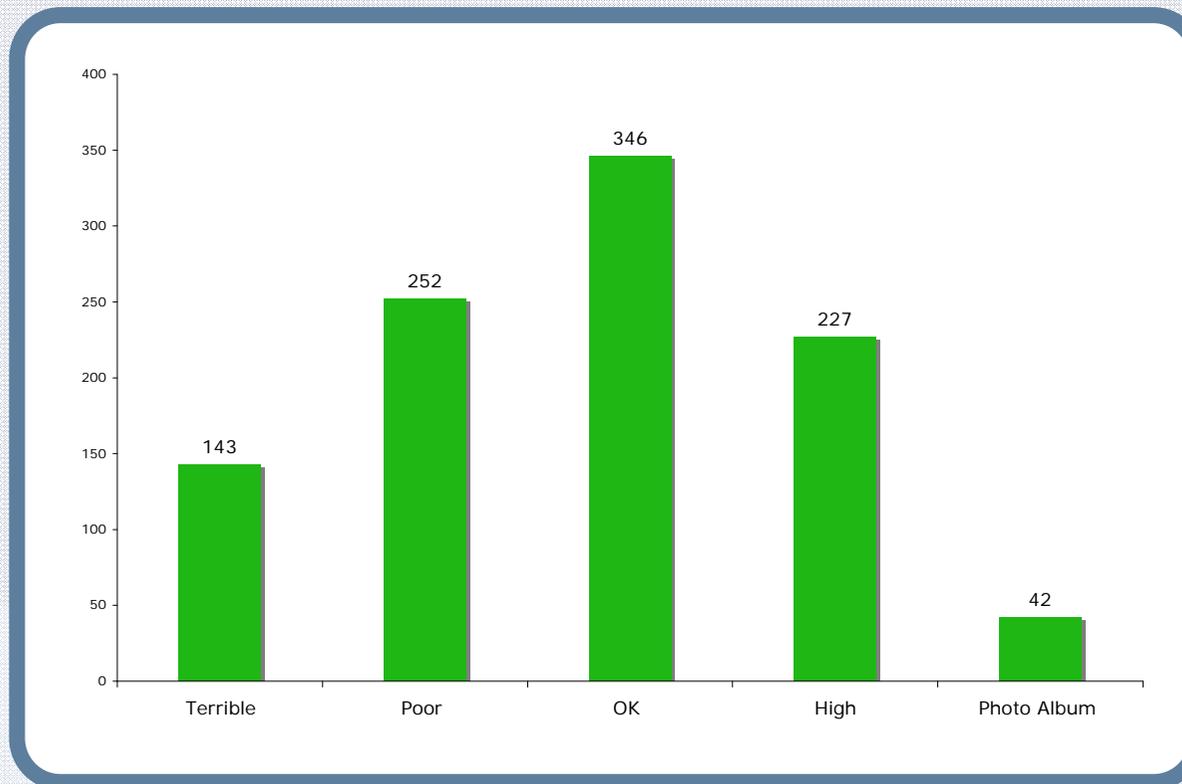
- **Randomly selected 1000 images from 1 million**
- **Manually annotated for quality**

Analysis: Comparison to Normal Photos

	HDM (1000 from 1M)	Personal Photos (10,523)
People	29.9	30.5
Buildings	3.5	35.0
Indoor	73.4	15.2
Outdoor	5.1	84.8
Cityscape	2.1	22
Landscape	1.1	23.8
Computer Screen/TV	26.7	0.0
Conversation Scenario	13.3	< 1.0
In a Vehicle	16.2	0.0
Work Scenario	24.6	0.0
Back view of People	6.8	2.0

Organising Visual HDMs

Event Detection

Integrating Context

Content Analysis

Suggested Organising Process

- **Event detection and segmentation**
  - To divide a HDM into a set of discrete events
  - Fixed events or dynamic events?
- **Visual feature extraction**
  - To add automatic semantic annotations to the content
- **Integration of context information**
  - Additional semantic enrichment
- **Human annotation**
  - Amend / add to the automatic semantic annotation and enrichment

Event Detection & Segmentation

- Segment 4,500 photos per day into a set of events (F 0.62 for visual features & sensors, 0.55 for visual alone)
  - Similar to SBD in digital video processing
  - We employ visual features and output of on-device sensors

Event Segmentation

Multiple Events

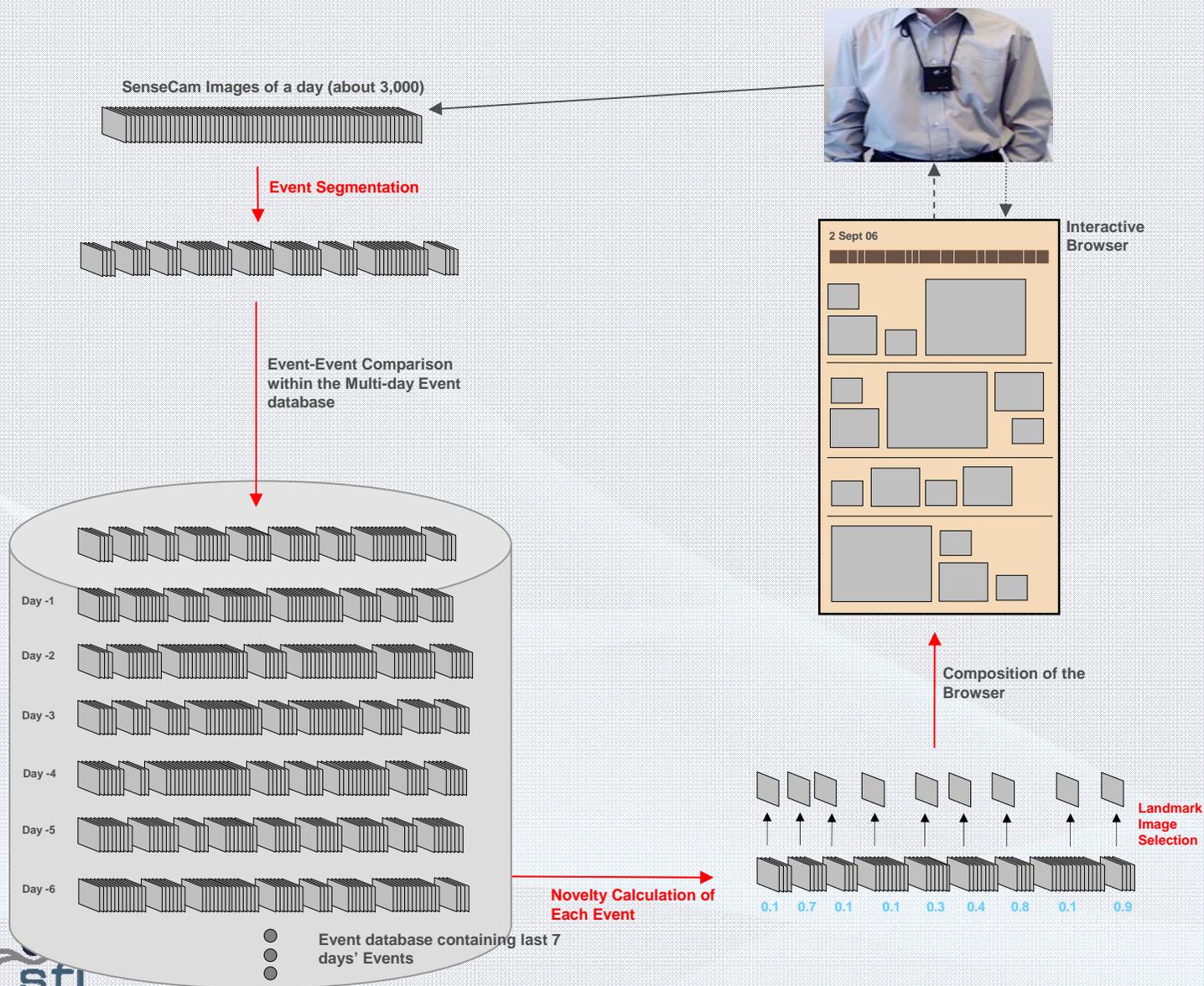| Finishing work in the lab | At the bus stop | Chatting at Skylon Hotel lobby | Moving to a room | Tea time | On the way back home |

Summarisation

Daily Browser Overview

SenseCam Images of a day (about 3,000)

Event Segmentation

Event-Event Comparison within the Multi-day Event database

Day -1
Day -2
Day -3
Day -4
Day -5
Day -6

Event database containing last 7 days' Events

Novelty Calculation of Each Event

0.1 0.7 0.1 0.1 0.3 0.4 0.8 0.1 0.9

Landmark Image Selection

2 Sept 06

Interactive Browser

Composition of the Browser

Cathal Gurrin (2008)

Integration of Context

- Help bridge the Semantic Gap
  - Shown to work for photo search

- We integrate:
  - GPS location
  - Social context, e.g. people near me,
  - Environmental context, e.g. temperature, light, acceleration
  - Biometric metadata
  - Communication interactions

Integration of Content Analysis

- We know that analysis content to identify visual features can work for both photo and video retrieval
  - May not accurate when examined independently, but useful within an overall content organisation scenario
- For HDMs, content analysis can:
  - Provide useful organisation functionality for a visual HDM
    - Quality detection of photos
    - Near duplicate detection
  - Support indexing and retrieval
  - Support linkage among events based on concept co-occurrence
- However a new suite of content analysis tools will need to be developed.

Concept Detection Tools

- Like conventional photos, visual concept detectors can be used for semantic enrichment
  - However, the concepts will be different

Concept	Frequency	Concept	Frequency
Vehicles (external view)	**0.93%**	Vegetation	1.5%
Road	1.95%	Screen	**20.51%**
Inside of vehicle	**4.33%**	Paper/Book (reading)	1.79%
Indoors	37.39%	Meeting	4.63%
Door	1.12%	Office	**15.06%**
Outdoors	6.37%	Food (eating)	4.15%
Building	3.58%	Hands	20.6%
Tree	1.85%	Holding a cup/glass	0.81%
Grass	0.86%	Faces	5.84%
Sky	2.93%	People	12.83%

Visual Feature Extraction

Suite of Sensecam specific concept detectors under development (with MediaMill):

- Steering wheel (72%)
- Shopping (75%)
- Inside of vehicle when not driving (airplane, taxi, car, bus) (60%)
- Toilet/Bathroom (58%)
- Giving Presentation / Teaching (29%)
- View of Horizon (23%)
- Door (62%)
- Staircase (48%)
- Hands (68%)
- Holding a cup/glass (35%)
- Holding a mobile phone (39%)

- Eating food (41%)
- Screen (computer/laptop/tv) (78%)
- Reading paper/book (58%)
- Meeting (34%)
- Road (47%)
- Vegetation (64%)
- Office Scene (72%)
- Faces (61%)
- People (45%)
- Grass (61%)
- Sky (79%)
- Tree (63%)

Other CDVP Sensecam Activities

Event Augmentation

Location Mapping

Event augmentation – Croke Park

Here's an image from a SenseCam after a Irish football game in Croke Park, Dublin. Can we see other people's pictures of this match.
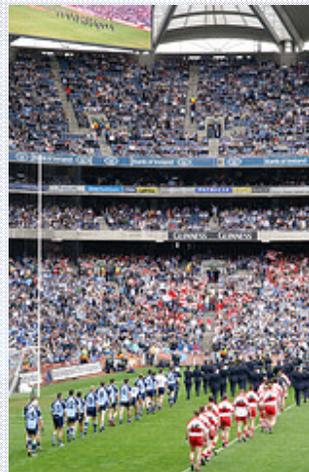
Let's search by location…

Event augmentation – Croke Park

- Receive the following pictures...
- Then filter out to just those results from the same day

Cathal Gurrin (2008)

Location Mapping Sensecam

An external GPS device maintains a log of where the wearer has been. Sensecam data is later location stamped by matching timestamps.

Location Mapping Sensecam

- Experimental integration of location stamped visual lifelog with Visual Mapping Software

- Typical Scenario:

  I recently visited Asia, find me a sequence of events where I was eating with other people in both Korea and China.

  - Requires:
    - GPS location stamping
    - Feature detectors for (people, food, eating, etc...)
    - Event Segmentation
      - With Key photo selection

Challenges for HDMS

Indexing Challenges
Retrieval Challenges

Indexing Challenges

- **Where to get the data?**
  - Very private data
- **To minimise the cognitive load on the user when uploading and indexing the content.**
- **To automatically organise the content into events which are meaningful units to the user.**
  - An indexing-time event segmentation is likely not to be sufficient
- **To automatically apply semantic enrichment techniques to the events and photos within the HDM, so as to provide easy and powerful retrieval.**

Retrieval Challenges

- How to evaluate?
  - The data evaluation is inherently subjective to the data gatherer
- To provide fast search facilities for millions of photos per year, many tens of millions of sensor data, and tens of thousands of human interactions.
- To automatically link between related events and provide a browsing/search mechanism.
- To provide high precision of retrieval.
- To support context sensitive retrieval, in that the HDM can seek out and locate events based on the current actions of the user.

Human Digital Memories
My thoughts…

Experiences
Thoughts
Who would do this?
Why?

Summary of a HDM
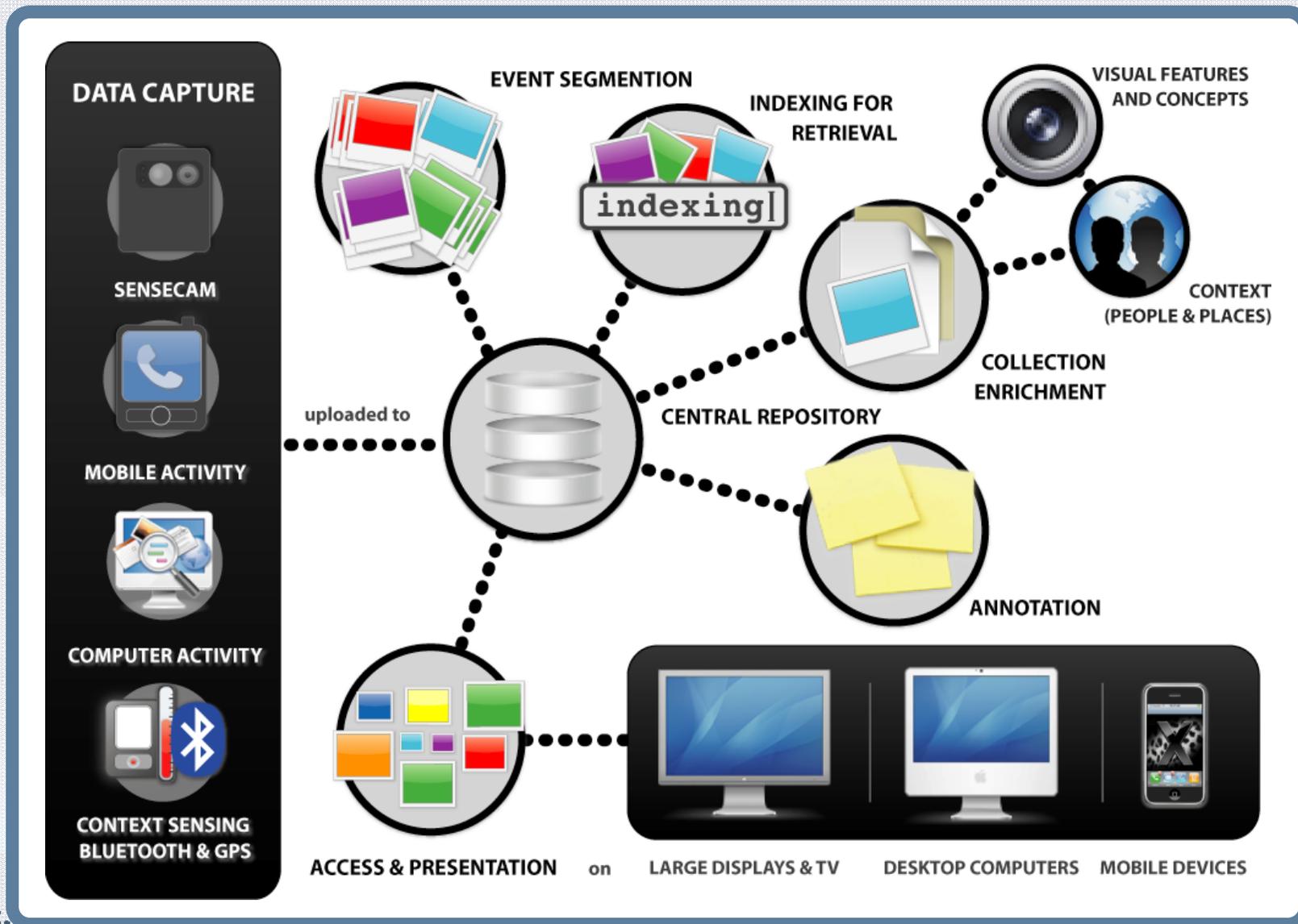
Experiences of wearing a Sensecam

- **I decided to become a long-term lifelogger back in mid 2006**
  - Over 2.5 million SenseCam images
  - Each with GPS position
- **Summary of my experiences:**
  - Most people don't notice the camera
    - Those that do always remember!
    - Most people don't mind the camera
  - About 40% of photos captured are low quality, more are stop-photos (boring photos of typical scenes like driving or working at desk).
  - Need extremely understanding family, girlfriend and friends

Millionth Image

Favourite Image

My thoughts after 16 months

- Importance of contextual semantic enrichment
- Event browsing is key
  - Too many photos to browse, need event summary and then 'drill down' to view event in detail if required
  - Stop events, (like work desk and driving) can be hidden.
- 'Total Recall', little sign of 'Event Decay'
  - I remember nearly every (non stop-) event when I see it…
- How do I want to interact with my HDM?
  - I want seamless integration with my context (recommendation)
  - I want powerful and fast search
    - Not only time/date, but location, people, biometrics
  - I want blog-style sharing of important events
    - Ideally directly from the device

Why would people do this?

- **Because we can:**
  - My life on 1TB of photos, audio another TB.
  - In 2020 2TB = €10
  - People blog, so this is an extreme extension of blogging
- **Health Benefits of Lifelogging**
  - The Frammington Study has shown the benefits of logging an aspect of peoples lives.
  - Much research into Sensecam wearing for Alzheimers sufferers.
- **Finally, is it a SnowBall Technology?**
  - Like facebook, my friends have one, so I need one too…

Cathal Gurrin (2008)

Who would use it?

- **People who want a digital memory of their life**
  - Automatic visual diaries, richer in data than we could imagine
  - Ability to show like experiences to others
- **People with memory difficulties**
  - Microsoft study has shown initial memory benefits
- **Bloggers keen for a new technology, more enriched blogs**
- **Medical professionals**
- **Security professionals**
  - UK police testing head mounted video cameras
- **Many more!**

Xie Xie

http://www.cdvp.dcu.ie

Acknowledgements:

Questions ?