Context and Content Analysis for Managing Personal Media Archives

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

Cathal Gurrin (2008)
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.

Sometimes it's for a reason, e.g. security personnel, medical staff, personal e.g. diaries, blogs, etc.

Sometimes it's for posterity, recording vacations, family gatherings, social occasions;

Sometimes it's 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
What comprises a HDM?

- Humans have been finding ways to substitute memory for a long time
  - Druids, books, computers…
  - MEMEX from ‘As We May Think’

- For HDMs we can…
  - Log our environment;
    - Traffic, pollution, **people near us**, etc…
  - Log our **location**
  - Log what we see and do
    - **Visually**, aurally,
  - Recording biometrics;
    - Log our health, fitness, activity, etc…
  - Log our actions
    - Communications
    - Data interactions
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
  - CDVP 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 CDVP Experience in Managing 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 (one type of context) has been used to support organisation (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
On the table

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 tourism guide, http://www.dublin-tour.ie

COMMENTS

[User profile]
Nice photo! I hope you guys enjoyed the event.
(7 Jun 2005)

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

[User profile]
Ah, thanks for uploading this photo!
(23 Jul 2005)

[User profile]
No problem, I had it for a long time but hadn't uploaded.
(24 Jul 2005)

[User profile]
I like the colour - well done!
(1 Aug 2005)

[User profile]
Ireland must be a good place for celebrating a
What we have learned?

- Event segmentation is important to organise the content
- 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 much easier to recall!
- 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
Video Search

- Video data organised into video shots and scenes for easier organisation
- Extensive use of Metadata for searching
- Also, research effort on 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
Cathal Gurrin (2008)

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 "DELETE" button to improve subsequent query.

RESULT PAGE: 1 2 3 4 5

1: ABC News (30 Oct 1998)

had failed repeatedly and publicly and threaten tests...

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

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

that manages about we’ve get big he’d be in an beard 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 lines 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.


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

the government the first person in space this all u.s. launch rockets had failed repeatedly and publicly and there were tests pressure was building so many american investments there are the past there have been
What have we learned?

- Shot Boundary Detection and scene detection is important to organise the video content into meaningful units.
- Content analysis tools can help:
  - 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 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

Cathal Gurrin (2008)
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
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.
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: Quality

- Randomly selected 1000 images from 1 million
- Manually annotated for quality

Cathal Gurrin (2008)
• 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: Comparison to Normal Photos

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

Cathal Gurrin (2008)
Organising Visual HDMs

Event Detection
Integrating Context
Content Analysis
Suggested Organising Process

1. Event detection and segmentation
   - To divide a HDM into a set of discrete events
   - Fixed events or dynamic events?

2. Integration of content analysis
   - To add automatic semantic annotations to the content

3. Integration of context information
   - Additional semantic enrichment

4. Indexing and presentation

5. Possible human annotation
   - Amend / add to the automatic semantic annotation and enrichment

Cathal Gurrin (2008)
1. 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)
  - We employ visual features and output of on-device sensors
2. 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

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

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%)
3. Integration of Context

- 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
4. Indexing and Presentation

DATA CAPTURE
- SENSECAM
- MOBILE ACTIVITY
- COMPUTER ACTIVITY
- CONTEXT SENSING BLUETOOTH & GPS

EVENT SEGMENTATION
INDEXING FOR RETRIEVAL
INDEXING FOR RETRIEVAL

uploaded to
CENTRAL REPOSITORY

VISUAL FEATURES AND CONCEPTS
CONTEXT (PEOPLE & PLACES)

COLLECTION ENRICHMENT
ANNOTATION

ACCESS & PRESENTATION on
LARGE DISPLAYS & TV
DESKTOP COMPUTERS
MOBILE DEVICES

Cathal Gurrin (2008)
4. Indexing and Presentation

- SenseCam Images of a day (about 4,500)
- Event Segmentation
- Event-Event Comparison within the Multi-day Event database
- Event database containing last 7 days’ Events
- Composition of the Browser
- Novelty Calculation of Each Event
- Landmark Image Selection

Cathal Gurrin (2008)
29 May 2006

Drag the slider bar to adjust the number of important events.

Cathal Gurrin (2008)

Similar Events
92 Similar Events have been found. Click on the photo to replay all photos within the Event.

Sort by: TIME | SIMILARITY | PEOPLE

11:16 (Duration: 14m 05s)
11 APR 2008

13:45 (Duration: 14m 05s)
11 APR 2008

10:31 (Duration: 23m 55s)
17 APR 2008

14:39 (Duration: 15m 33s)
12 APR 2008

09:52 (Duration: 05m 23s)
12 APR 2008

18:19 (Duration: 21m 11s)
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
• 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 & Retrieval Challenges
Challenges for HDMs

- Where to get the data?
  - Very private data, not easily available
- How to evaluate?
  - The data evaluation is inherently subjective to the data gatherer
- Need to:
  - minimise the cognitive load during indexing.
  - organise into events (dynamic?).
  - automatically apply semantic enrichment techniques
  - provide fast search facilities for millions of photos per year, many tens of millions of sensor data, and tens of thousands of human interactions.
  - automatically link between related events
  - provide high precision of retrieval.
  - support context sensitive retrieval
Human Digital Memories
My thoughts…

Experiences
Thoughts
Who would do this?
Why?
Experiences of wearing a Sensecam

• I have been gathering a HDM (visual & location) since June 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

Cathal Gurrin (2008)
Some comments...

- 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.

• Who would do this?
  – Security/Health professionals
  – people with memory difficulties
  – Bloggers & people who want an automatic diary of their life
Xie Xie
http://www.cdvp.dcu.ie

Acknowledgements:

Microsoft
Irish Research Council for Science Engineering and Technology
Science Foundation Ireland (grant 07/CE/I1147)

Questions ?