Life-logging: what’s it about?

Petia Radeva

www.cvc.uab.es/~petia

Barcelona Perceptual Computing Laboratory (BCNPCL), Universitat de Barcelona (www.bcnpcl.wordpress.com) &

Computer Vision Center (www.cvc.uab.es)
The life-logging trend

Life-logging and egocentric vision

- Video segmentation for events extraction

- Motion-based video segmentation towards activities recognition

- Human tracking, towards social interaction and key-frame extraction

- Active learning for object recognition

- Object discovery for lifestyle characterization
Wearable cameras and the life-logging trend

The 7 Best Inventions Of 2012

The Huffington Post | By Betsy Isaacson
Posted: 12/28/2012 10:49 am EST | Updated: 12/28/2012 1:57 pm EST

Shipments of wearable cameras worldwide from 2013 to 2015 (in millions)

Forbes

2014 Will Be The Year Of Wearable Technology
**Definition**: Life-logging consists of acquiring images related to an individual through a wearable camera.

**Benefits:**

- **A digital memory** of people you met, conversations you had, places you visited, and events you participated in.
  - This memory would be searchable, retrievable, and shareable.

- **A 14/7/365 monitoring of daily activities**.
  - This data could serve as a warning system and also as a personal base upon which to diagnosis illness and to prescribe medicines.

- **A way of organizing, shaping, and “reading” your own life**.
  - A complete archive of your work and play, and your work habits. Deep comparative analysis of your activities could assist your productivity, creativity, and consumptivity.

- To the degree this life-log is shared, this archive of information can be leveraged to help others work, amplify social interactions, and in the biological realm, shared medical logs could rapidly advance medicine discoveries.
Technology is “running”!

Evolution of life-logging apparatus, including wearable computer, camera, and viewfinder with wireless Internet connection. Early apparatus used separate transmitting and receiving antennas. Later apparatus evolved toward the appearance of ordinary eyeglasses in the late 1980s and early 1990s.

“Quantified Self & life-logging Meets Internet of Things (IOT)” by Mazlan Abbas.
Is it feasible to record everything that happens in a person’s life?

The Moore’s Law (1965): “Transistor density that can be etched onto the silicon wafer of a microchip doubles every two years”.

- In 1970, a disk to store 20 MB was the size of a washing machine and costed 20,000$.
- Today a TB (one trillion bytes) costs a 100$ and is the size of a paperback book.
- By 2020 a TB will cost the same as a good cup of coffee and will probably be in your cell phone.
- 100$ will then buy around 250 TB of storage, enough to hold tens of thousands of hours of video and tens of millions of photographs.
- This should satisfy most life-loggers’ recording needs for an entire life.
- In fact, digital storage capacity is increasing faster than our ability to pull information back out.
- From 2000 it became trivial and cheap to sock away tremendous piles of data.

The hard part is no longer deciding what to hold on to, but how to efficiently organize it, sort it, access it, and find patterns and meaning in it.

This is a primary challenge for the engineers that will fully unleash the power of Total Recall.”

Will life-logging and internet of things help know us better?

Don’t you love to know...

- Where you’re going?! Who you’ve interacted with?!
- How long you’ve spoken to friends?! The affinity to connections?!
- How long it takes to get to work?!
- The tone of your messages?! The amount you text, tweet, or update?!
- How much exercise you’re getting?!
- How much you get distracted?! Where is your time most spent?

Life-logging everything you see, do, feel, speak, experience and hear is almost here.

The really big issue here is that you might, individually, not worry about publishing details of your personal life.

• But you are publishing your friends, family and business contacts details at the same time.

Scared?
Ethical guidelines for wearable cameras

- **Anonimity and confidentiality**: Researchers coding image data should:
  - not discuss the content with anyone outside of the team,
  - not identify anyone they recognize in the images,
  - be aware of how sensitive the data are.

- **Data encryption**: Confidentiality can be protected by configuring devices and using specialist viewing software to make the images accessible only to the research team (lost devices).
  - Devices should be configured so that data can only be retrieved by the research team. It should be impossible for participants or third parties who find devices to access the images.

- **Data storage**: Collected images should be stored securely and password-protected, according to national regulations.
Understanding user privacy requirements and risks from emerging technologies

Wearable cameras can be very useful:

An estimated one million Russian motorists have dashboard video cameras installed in their cars.

Police officers carrying video camera units and using Velcro to place these cameras in police wagons, helmet cams, ear cams, chest cams with audio capability, GPS locators, taser cams

• “Even with only half of the 54 uniformed patrol officers wearing cameras at any given time, a department in USA had an 88 %decline in the number of complaints filed against officers, compared with the 12 months before the study”, The New York Times, 4th of July, 2013.

People are bad at:

1. understanding the future value of revealing private information today,
2. understanding the risks from technology they have not yet used or heard of.
Benefits and potential applications

- It will take quite some time for people to feel comfortable with ‘always connected’ devices that can discreetly take photos or videos. Will the benefits outweigh the negatives?

- Wearable camera can provide many benefits, such as assistive technologies to help people:
  - see better,
  - store and remember better,
  - work better,
  - function better,
  - remember and recognize names and faces, etc, etc

“Quantified Self & life-logging Meets Internet of Things (IOT)”, Dr. Mazlan Abbas, MIMOS Berhad
The life-logging trend

Egocentric vision for life-logging

Video segmentation for events extraction

Motion-based video segmentation towards activities recognition

Human tracking, towards social interaction and key-frame extraction

Active learning for object recognition

Object discovery for lifestyle characterization
"Taking photos can impair your ability to remember" (1)

Next time you're at a museum or an event, think before you snap a photo!

"SenseCam has already made an impact in memory enhancement" (2)

What do we have?

- Self centric, automatically captured Images of our life
- Objective photos versus Subjective (traditional) photos
- Most similar photos to our memory
Motivation

“Taking photos can impair your ability to remember” (1)

Next time you're at a museum or an event, think before you snap a photo!

“SenseCam has already made an impact in memory enhancement” (2)

What do we have?

- Self centric, automatically captured Images of our life
- **Objective** photos versus **Subjective** (traditional) photos
- Most similar photos to our memory

1) University of Groningen, November, 2014
Life-logging data

What we want:

Events to be extracted from life-logging images
Life-logging data

What we have:
We propose an energy-based approach for motion-based event segmentation of life-logging sequences of low temporal resolution. The segmentation is reached integrating different kinds of image features and classifiers into a graph-cut framework to assure consistent sequence treatment.

We chose a SenseCam or Narrative cameras hung on the neck or pinned on the dress that capture 2–4 fps. Complete dataset of a day captured with SenseCam (more than 4,100 images).

Choice of devise depends on: 1) where they are set: a hung up camera has the advantage that is considered more unobtrusive for the user, or 2) their temporal resolution: a camera with a low fps will capture less motion information, but we will need to process less data.

Or the hell of life-logging data

100,000 images per month
Towards lifestyle characterization

We want to extract life-logging information about:

- Events
- Activities
- Social interactions, etc.
- Memorable moments
- Personal habits and context
- Lifestyle... healthy lifestyle...

Our egocentric vision research:

- Video segmentation for events extraction
- Motion-based video segmentation towards activities recognition
- Human tracking, towards social interaction and key-frame extraction
- Active learning for object recognition
- Object discovery for lifestyle characterization, etc, etc.

What? Where? When? Who?
Egocentric vision

- Video segmentation for events extraction
- Motion-based video segmentation towards activities recognition
- Human tracking, towards social interaction and key-frame extraction
- Active learning for object recognition
- Object discovery for lifestyle characterization
Life-loggin (LL) devices are characterized by easily collecting huge amount of images.

One of the challenges of life-loggin is how to organize the big amount of image data acquired in semantically meaningful segments in order to be able to store them and review later, being able to focus just on the most important aspects.
Methodology

Data set: thousands of images

Features Extraction
- RGB+HOG
- CNN

Clustering
- Ward
- k-Means
- Spectral Clustering

Daily Video Summarization

Event extraction

Event characterization
Methodology

- Features
  - RGB + HOG

- Convolutional Neural Networks (4096 Features)

Features:
- Color
- Structure

Convolutional Neural Networks:
- 4096 Features

University of Groningen, November, 2014
Clustering Techniques

- **Spectral Clustering**
- **K-Means**
- **WARD**

Image shows a dendrogram with data points clustered into different groups, illustrating the WARD linkage method. The eigenvalues are represented by a matrix, indicating the first three eigenvectors.
Validation

Results

Qualitative
Validation

Results

Moments of change
Validation

Results

Quantitatives

\[ JC = \frac{|C_{\text{auto}} \cap C_{\text{man}}|}{|C_{\text{auto}} \cup C_{\text{man}}|} \]

Clustering RGB+HOG features

Clustering CNN features

Weighted Features

<table>
<thead>
<tr>
<th>Feature</th>
<th>K-Means</th>
<th>Ward</th>
<th>Spectral Cluster</th>
</tr>
</thead>
<tbody>
<tr>
<td>CNN</td>
<td>56</td>
<td>60</td>
<td>50</td>
</tr>
<tr>
<td>RGB+HOG</td>
<td>35</td>
<td>37</td>
<td>42</td>
</tr>
</tbody>
</table>

University of Groningen, November, 2014
Event segmentation in LL data is characterized by the movement of the person wearing the device.

Group consecutive frames in three general event classes:

- "Static" (person is not moving)
- "In Transit" (person is moving or running)
- "Moving Camera" (person is in the same area performing some action).
Egocentric vision

- Video segmentation for events extraction
- Motion-based video segmentation towards activities recognition
- Human tracking for social analysis
- Active learning for object recognition
- Object discovery for lifestyle characterization
Our goal on social analysis

- **Who** was with us? For **how long**?
  - **Humans** are important

- **Video segmentation** based on presence of **people**
  - Each segment is an event related to specific person

- Intuitive solution is to **track visible people**
What does make the tracking even more difficult on life-logging data?

- Low spatial and temporal resolution of images
- Free motion of the camera
- Frequent scene occlusion and distortion
- Wide variation in appearance, scale and location of people along the videos

Even best state-of-the-art tracking methods fail!
Our proposal: Bag-of-Tracklets

Our strategy: Study restrictions of current tools to design a new method

A tracking method independent from background changes and temporal resolution could fit this data

What we propose:

- Treat every detection individually
- Track every detection (as much as we can!)
- Try to find one reliable track of a same person, among many (one tracklet per detection) by grouping similar tracklets.

Advantage: Confront the tracking problem from higher level of information

- Tracklets instead of detection
- Getting rid of false alarms by excluding unreliable tracklets using group of information (bag-of-tracklets)
- Accuracy and robustness improved!

University of Groningen, November, 2014
Our Proposed Method

A day’s SenseCam images (3,000 – 4,000)

Video

- Seed Generation
- Tracklet Creation
- Bag-of-tracklets Formation
- Density Calculation
- Finding Reliable BOTs
- Prototype Extraction

Summarized Video

Person 1  Person 2  Person 3
Seed Generation

Video

Seed Generation
Tracklet Creation
Bag-of-tracklets Formation
Density Calculation
Finding Reliable BOTs
Prototype Extraction

Summarized Video
Tracklet Creation

- Video
  - Seed Generation
  - Tracklet Creation
  - Bag-of-tracklets Formation
  - Density Calculation
  - Finding Reliable BOTs
  - Prototype Extraction

Backward CT  Forward CT

Summarized Video

Tracklet generated by compressive tracking
Similarity btw tracklets:

\[ L(t^i_k, t^i_k) = \begin{cases} \frac{\text{box}(t^i_k) \cap \text{box}(t^i_k)}{\text{box}(t^i_k) \cup \text{box}(t^i_k)}, & \text{if } \text{box}(t^i_k) \neq \emptyset \text{ and } \text{box}(t^i_k) \neq \emptyset \\ 0, & \text{otherwise.} \end{cases} \]

Likelihood that a tracklet belongs to a model:

\[ L(t^j, M) = \frac{1}{n} \sum_{i=m_1, m_2, \ldots, m_n} \frac{1}{|t^i_m|} \sum_{k=1}^{t^i_m} L(t^i_k, t^i_k) \]
Bag-of-Tracklets Formation

1. Video
2. Seed Generation
3. Tracklet Creation
4. Bag-of-tracklets Formation
5. Density Calculation
6. Finding Reliable BOTs
7. Prototype Extraction

Summarized Video

University of Groningen, November, 2014
Density Calculation

\[ d(B) = \frac{N_B}{\text{median}_{t^i \in B}(\text{length}(t^i))} \]

- \( N_B \) = number of tracklets in the bag \( B \)
- \( t^i \) = tracklet \( i \) in the bag \( B \)
Finding Reliable BOTs

1. Video
2. Seed Generation
3. Tracklet Creation
4. Bag-of-tracklets Formation
5. Density Calculation
6. Finding Reliable BOTs
7. Prototype Extraction

Summarized Video

University of Groningen, November, 2014
Prototype Extraction

- Video
- Seed Generation
- Tracklet Creation
- Bag-of-tracklets Formation
- Density Calculation
- Finding Reliable BOTs
- Prototype Extraction

Summarized Video
Results:

Dataset:

<table>
<thead>
<tr>
<th># Days</th>
<th># Frames</th>
<th># Frames with Person(s)</th>
<th># Trackable Person</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>24,000</td>
<td>11,000</td>
<td>65</td>
</tr>
</tbody>
</table>

Results:

- Accuracy: A model similar to Jaccard Distance
- # True Positive tracked frames / # Frames for that track in the groundtruth
- Average **accuracy of 84%** has been obtained

A sequence of 10 frames, with detected people indicated

Tracking results using Bag-of-Tracklets
Comparing to the performance of Compressive Tracking (without using BOT) over SenseCam images [48%], **75% of improvement** has been achieved!

False alarms in detection and compressive tracking

**Tracklet A**

**Tracklet B**

True track of the person using Bag-of-Tracklets
Advantages of Bag-of-Tracklets approach

Advantages

- Efficient method to track persons in low spatial and temporal resolution images
- Human-based segmentation approach of visual life-logging data
- Detect events based on human presence
- Address the tracking problem using higher level of information
- Assignment problem

Limitations and current work

- Occlusion
- Cost
- Pose recovery
Egocentric vision

- Video segmentation for events extraction
- Motion-based video segmentation towards activities recognition
- Human tracking, towards social interaction and key-frame extraction
- Active learning for object recognition towards life characterization
- Object discovery for lifestyle characterization
Towards lifestyle characterization

Our next steps are directed towards visualizing summarized lifestyle data to ease the management of the user’s healthy habits (sedentary lifestyles, nutritional activity of obese people, etc.).

Life-logging can help us accomplish our goals: taking photos of our everyday life and being able to analyse what we eat, starting by the dish recognition.

**We need a food-related objects classifier!**
But, how can we automatically detect every instance of a dish in all of its variants, shapes and positions and in such a large number of images?

The main problems that arise are:
• Complexity and variability of the data.
• Huge amounts of data to analyse (up to 100,000 images per month).

Any efficient supervised classifier needs a huge amount of training set!

Using human time for labelling million of images is too expensive.

We need an automatic aid for labelling huge amount of images.
Goal: Minimize human effort by guiding the process of creating the training set.

Effect of active learning: improvement of learning performance (left), improvement of training time (right).
Data sets

- Images from about 15 days
- Each day has up to 4,500 pictures.
- Total: 43,750 images.
- 2,900 images per day on average.

SenseCam images

ImageNet images
Active learning for nutrition-related objects annotation

Active forest reduces by 70% the number of clicks per label training set.

Egocentric vision

- Video segmentation for events extraction
- Motion-based video segmentation towards activities recognition
- Human tracking, towards social interaction and key-frame extraction
- Active learning for object recognition
- Object discovery for lifestyle characterization
What is the context of a person day-by-day?

Do these sets belong to the same person?


University of Groningen, November, 2014
Main Goal

Goal: to develop automatic techniques for object discovery to characterize the environment of the person wearing the camera.

Our proposal: to discover iteratively the most relevant objects from egocentric videos for a particular user, based on:

1. Clustering of object region candidates
2. CNN feature extraction
3. New refill methodology on already discovered instances clusters.

Images acquired by a life-logging device, where objects of interest appear like: mobile phone, person, or TV monitor.

University of Groningen, November, 2014
The selected easiest samples are complemented with a certain percentage (20%) of samples from the previous knowledge. Equally divided between all the known classes but the "NoObject".

Before starting with the discovery iterations, 40% of all the object candidates are inserted into the "previous knowledge" set.

Algorithm

Used a pre-trained CNN provided by Hinton et al., trained on millions of ImageNet images in a succession of convolutional and pooling layers. Deleted the last layer (supervised) and used the output of the penultimate layer as our features (4096 variables).

University of Groningen, November, 2014
Our dataset consists of 1,000 images from a person's work day, from which 50,000 object candidates were extracted. To validate our method, we used the labels of the most frequent objects appearing.
As expected, more than 76% of the samples were labeled as “No Objects”.

We defined three different test settings to evaluate our proposal:

1. CNN Features
2. CNN Features with the Refill methodology
3. Features of [13] (Lee and Grauman’s work)
Goal: to develop tools for memory refocusing of MCI and Alzheimer people.

To develop, for subjects with MCI, a program-based life-logging captured by a Wearable Camera recording specific autobiographical episodes for stimulating posteriorly episodic memory function known to be deficient in MCI.

To explore the association between biomarkers changes in cognitive, functional and emotional outcomes.

To learn more about the underlying biological mechanisms for how effective behavioural interventions improve cognitive and functional outcomes.
life-logging for welfare

To derive lifestyle patterns from visual life-logs and to conduct a study on the feasibility of automatically generation of lifestyle patterns and interpretations to be used in the future to improve lifestyle of individuals.

- how to extract semantic units related to the lifestyle and their context relation,
- how to segment life-log data into meaningful events,
- what are the semantic units that characterize the lifestyle of individuals,
- what is their relation and how the context affects them,
- how to extract and characterize lifestyles patterns,
- what is the healthstyle, etc.

University of Groningen, November, 2014
Life-logging is a very recent trend growing very fast and with a huge potential.

Technology for data acquisition and storage is ready.

Life-loggers are waiting for tools of egocentric vision to process the huge amount of data.

Algorithms urgently needed for:
- Shot boundary detection
- Scene segmentation
- Event detection
- Data mining
- Object and Activity recognition
- Video annotation
- Key frame extraction, memorability and esthetics
- Video summarization and browsing
- Query and retrieval
- Numerous applications to health & wellbeing, memory, safety, leisure, etc., etc.
Thank you for your attention!
How else can be LL useful?

The Economic Impact of Bad Meetings
IS CLEARING YOUR SCHEDULE BETTER FOR BUSINESS?

TED speakers David Grady and Jason Fried want to take meetings to task—and for good reason. Recent data shows that meetings can cost companies valuable time and money. From inefficiency to unseemly costs, are meetings really benefiting your organisation?

How much time do meetings waste?

Most employees attend an average of 42 meetings per month. Executives on average spend 40–50% of their working hours in meetings. Executives average 99 hours per week in meetings. Where 7 of those hours are unnecessary and poorly run, which is equal to over 2 months per year wasted.

How much money do meetings cost?

There are 3 billion meetings per year in the United States. There are nearly 11 million forms per day in the United States. Estimated to waste $37 billion a year in the U.S. A Fortune 50 company estimates losses in excess of $75 million per year due to poor meetings. A meeting between several managers or execs may cost upwards of $1,000 per hour in salary costs alone.