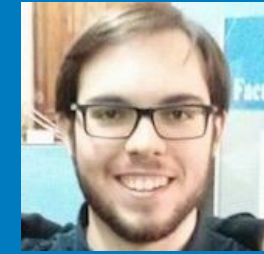# Semantic Summarization of Egocentric Photo Stream Events

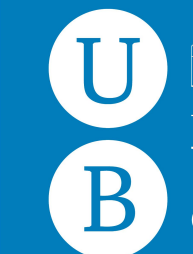Aniol Lidon Marc Bolaños Mariella Dimiccolli Petia Radeva Maite Garolera Xavier Giró-i-Nieto

UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH

Universitat de Barcelona

Motivation

Reminiscence therapy for dementia patients (44.4 million people worldwide in 2013).

Egocentric Photo Streams from a wearable camera (up to 2,000-3,000 pictures per day).

Content-based semantic summarization of events.
<u>Challenge</u>: Trade off Relevance vs Diversity
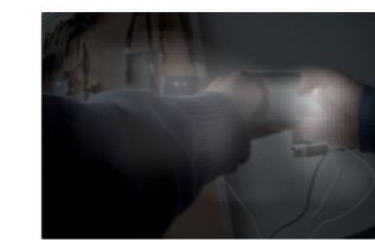
DIVERSITY
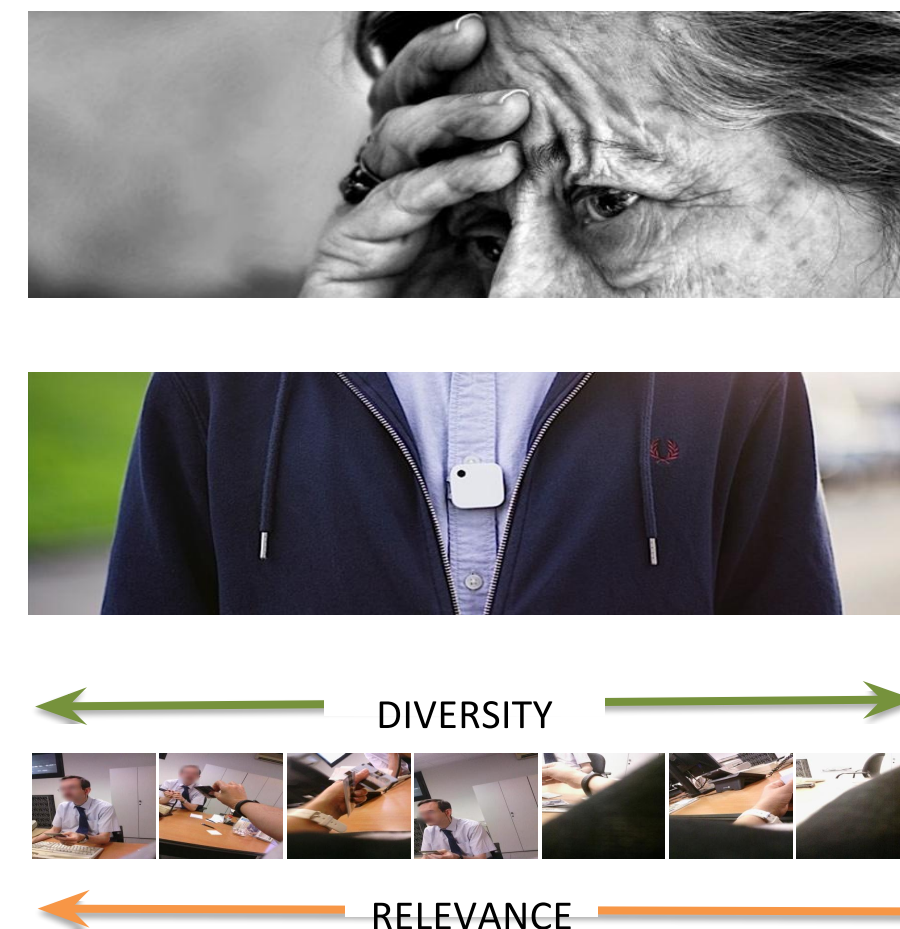
RELEVANCE
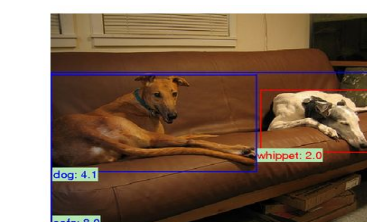
Relevance -aware Ranking

A ranked list $r_k(x)$ is built for each of the k semantic-aware rankings of M elements, with scores normalized by position. A weighted linear combination of scores to build r(x).

$$r_k(x) = \frac{M - R_k(x)}{M - 1}$$

$$r(x) = \sum_{k=1}^{3} w(k)\, r_k(x)$$

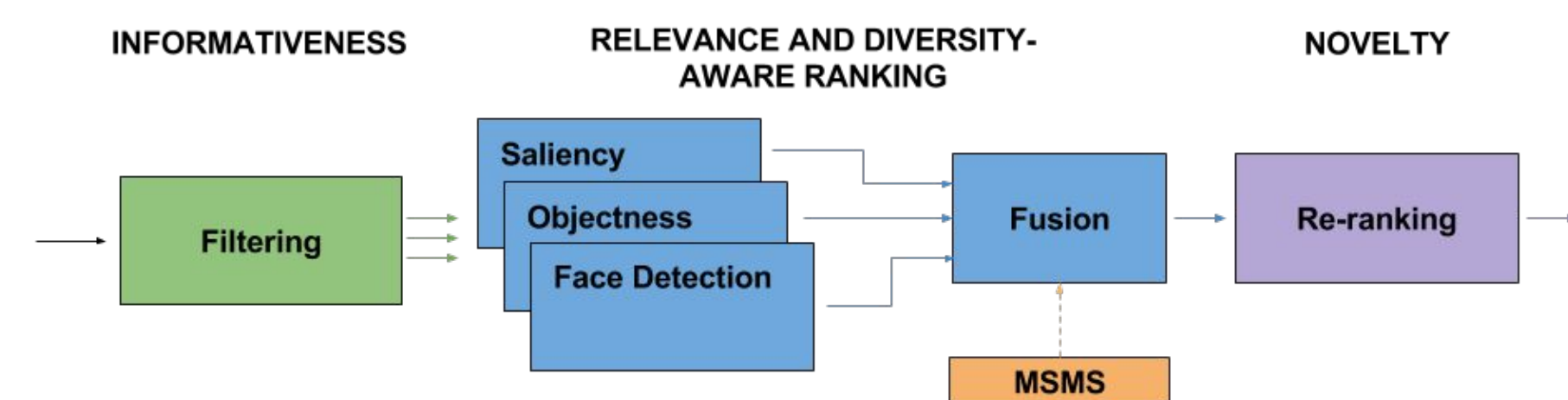In our work, we consider 3 semantic cues:

Saliency   Object Detection   Face Detection

System Architecture

INFORMATIVENESS   RELEVANCE AND DIVERSITY-AWARE RANKING   NOVELTY

Filtering → Saliency / Objectness / Face Detection → Fusion → Re-ranking

MSMS

Informativeness Filtering

CNN (CaffeNet) trained for classification between no-/informative images.
Non informative cases: blurry, sky, ceiling, wall, large occlusions...
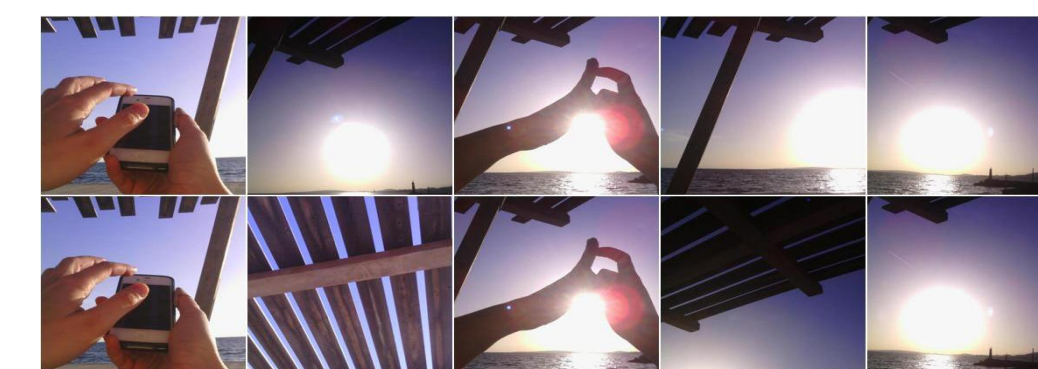
Informative

Non-Informative

Novelty-based Re-ranking
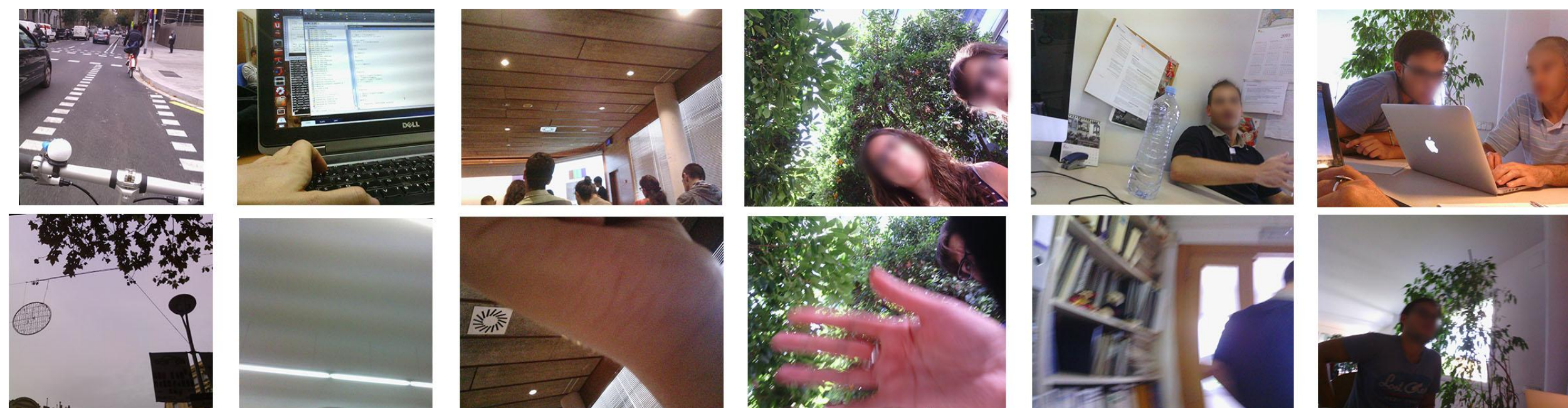
Greedy selection between candidate images x* to build set Y of t<M images leveraging relevance r(x) and novelty n(x,Y)....

$$x_{y_{t+1}} = \arg\max_{x^* \in \mathcal{X} \setminus \mathcal{Y}^t} (r(x^*) + n(x^*, \mathcal{Y}^t))$$

$$\mathcal{Y}^{t+1} = \mathcal{Y}^t \cup \{x_{y_{t+1}}\}$$

...where novelty is based on the visual similarity s(x,y) of a candidate image x* and the set Y.

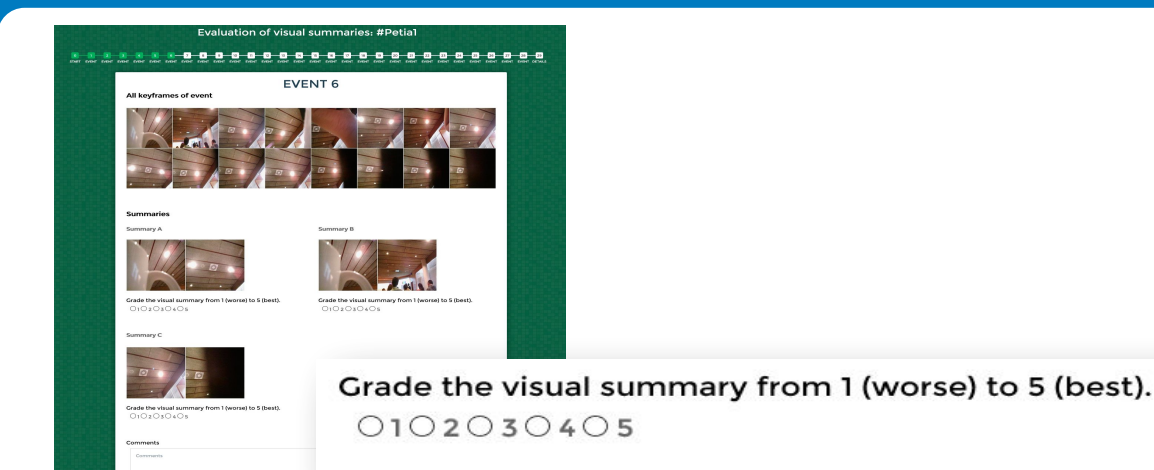$$n(x^*, \mathcal{Y}^t) = 1 - s(x^*, \mathcal{Y}^t)$$

Without Re-ranking

With Re-ranking

Evaluation with Mean Opinion Score

Grade the visual summary from 1 (worse) to 5 (best).

Our Solution	Ground-truth	Uniform samples
4.57	4.94	3.99

Caffe

Generalitat de Catalunya

NVIDIA.