Novelty Detection from Wearable Cameras

Omid Aghazadeh, Stefan Carlsson, Josephine Sullivan

Computer Vision Group, KTH

Stockholm, Sweden

June 2012

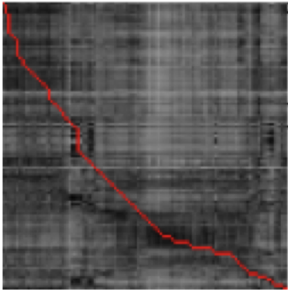
Our captured dataset

Deviations from repeated daily activity (manual labelling)

Register Query sequence to a Reference sequence

Given a query sequence and a reference sequence:

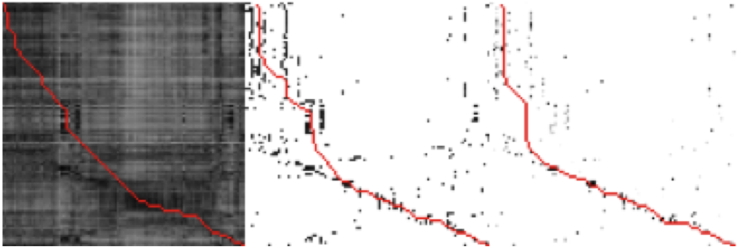
1. For each frame in the query sequence compute its similarity to all the frames in the reference frame.
2. Use a score based on *Bag Of Features* model.
3. Then use **dynamic time warping** to find the best temporal registration between the two sequences.



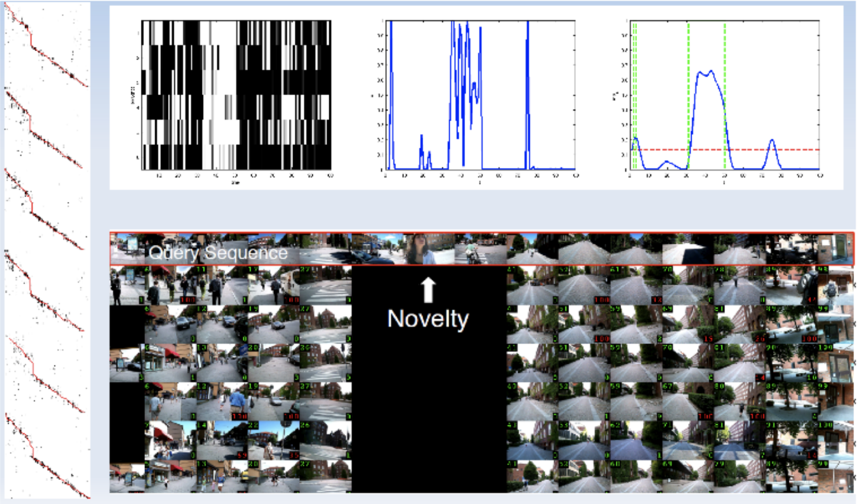
red curve is the minimum-cost temporal alignment

Register Query sequence to a Reference sequence

1. BOF similarity score is not sufficiently strong to produce good results.
2. Better results if one uses a similarity score using inliers from epipolar matching. But this is too expensive.
3. Apply it to the kNN of each query frame in the reference sequences using BOF model.

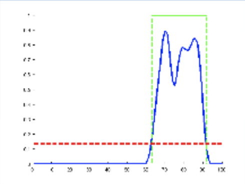


Match costs and novelty detection



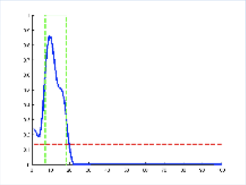
Novelty detection result

- Minimum match energy(top)
- Detected novelties (bottom)



Novelty detection result

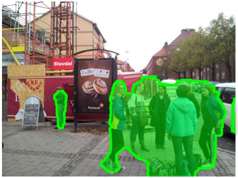
- Minimum match energy(top)
- Detected novelties (bottom)



Novelty - Background separation

I've been
here before!

What's new?



Register a reference and a query image using SIFT Flow

Query image: I_q

Reference image: I_r

Warped Reference image: $I_{r \to q}$

Compare registered images: **Color**

I_q

$I_{r \to q}$

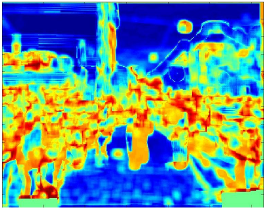
$I^{\text{err}}_{q,r,x} = \|I_{q,x} - I_{r \to q,x}\|$

$I_{i,x}$ is the color of pixel x in image i.

Compare registered images: **SIFT**

I_q

$I_{r \to q}$



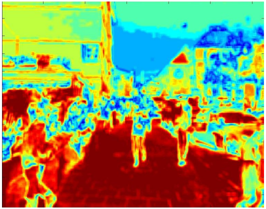
$S^{\text{err}}_{q,r,x} = \|S_{q,x} - S_{r \to q,x}\|$

$S_{i,x}$ is the SIFT feature vector computed at pixel x in image i.

Compare registered images: **NBS**

I_q

$I_{r \to q}$



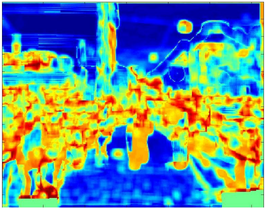
$$H^{\text{err}}_{q,r,x} = \sum_c QC(H_{q,c,x}, H_{r \to q,c,x})$$

- $H_{i,c,x}$ is the histogram of channel c intensity values from a patch centred at pixel x in image i.

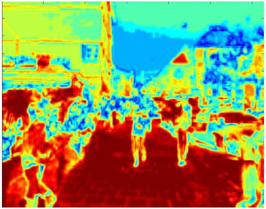
- $QC(\cdot,\cdot)$ is the Quadratic Chi distance between two histograms.

Compare registered images

$I^{\text{err}}_{q,r,x}$



$S^{\text{err}}_{q,r,x}$



$H^{\text{err}}_{q,r,x}$

For each pixel x in I_q (when compared with I_r) stack these features to get

$$F_{q,r,x} = \begin{pmatrix} I^{\text{err}}_{q,r,x} \\ S^{\text{err}}_{q,r,x} \\ H^{\text{err}}_{q,r,x} \end{pmatrix}$$

Compare Query Image to **all** Registered Images

- Have a set of reference images $\{I_r\}_{r=1}^{R}$.
- For each $I_{r \to q}$ compute $F_{q,r,x}$.
- At pixel x in I_q, aggregate the responses across different reference images by computing the
  1. algebraic mean
  2. harmonic mean
  3. minimum

  of the responses, to obtain the feature vector $F_{q,x}$

Probability of Novelty

- Use Logistic Regression to learn a mapping

$$g : \mathbb{R}^d \to [0, 1] \quad \text{s.t.} \quad g(F_{q,x}) \approx P(x \text{ is a novel pixel} \,|\, I_q, \{I_r\})$$

- Mapping learnt from manually annotated ground truth data.

Probability of Novelty thresholded

Compute query image feature descriptors

For every pixel x in the image extract a feature f_x based on its

- colour
- SIFT descriptor (PCA to 3 dims)
- position

Improve the segmentation via Graph Cuts

- Given an initial segmentation $l^{(0)} = \{l_x^{(0)}\}$

use **kernel density estimation** to estimate $p(f_x|l_x)$ as $\hat{p}_0(f_x|l_x)$ for both the background and foreground cases.

Improve the segmentation via Graph Cuts

- Use graph cuts to optimize

$$E(l) = \sum_{x \in \mathcal{X}} D_x(l_x) + \lambda \sum_{(x,y) \in \mathcal{N}} V_{x,y}(l_x, l_y)$$

where

$$D_x(l_x) = -\log(\,\hat{p}_0(f_x|l_x)\,) - \log(\,P(l_x)\,) + \log Z_x$$

and

$$V_{x,y}(l_x, l_y) = \begin{cases} \frac{1}{\|x-y\|} e^{-(I_B(x)-I_B(y))^2/(2\sigma^2)} & \text{if } l_x \neq l_y \\ 0 & \text{if } l_x = l_y \end{cases}$$

Iterate until convergence

1. Perform the maximization step

$$l^{(i+1)} = \arg\max_{l} \sum_{x \in \mathcal{X}} D_x(l_x) + \lambda \sum_{(x,y) \in \mathcal{N}} V_{x,y}(l_x, l_y)$$

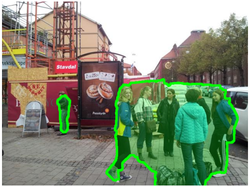
where

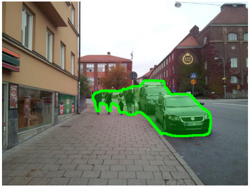
$$D_x(l_x) = -\log(\, \hat{p}_i(f_x|l_x)\,) - \log(\, P(l_x)\,) + \log Z_x$$

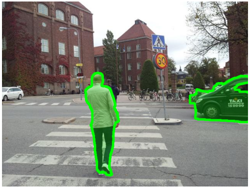
2. Update the kernel density estimates $\hat{p}_{i+1}(f_x|l_x)$ using the new labelling $l^{(i+1)}$.

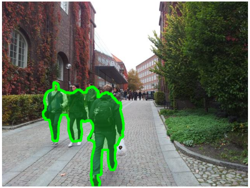

JÄRNVÄGSGRILLEN
7 ELEVEN

Final Segmentation









Final Segmentation