Novelty Detection from an Ego-Centric Perspective

Omid Aghazadeh, Josephine Sullivan, and Stefan Carlsson
Presented by Randall Smith

Outline

- Introduction
- Sequence Alignment
- Appearance Based Cues
- Geometric Similarity
- Example
- Dynamic Time Warping
- Algorithm
- Evaluation of Similarity Matching
- Results
- Conclusion

Introduction

- Problem: *Select relevant visual input from worn, mobile camera.*

- Motivation:

    - Routine Recognition [Blanke & Schiele 2009]

    - Life Logging [Doherty & Smeaton 2010]

      [Schiele et. al. 2007]

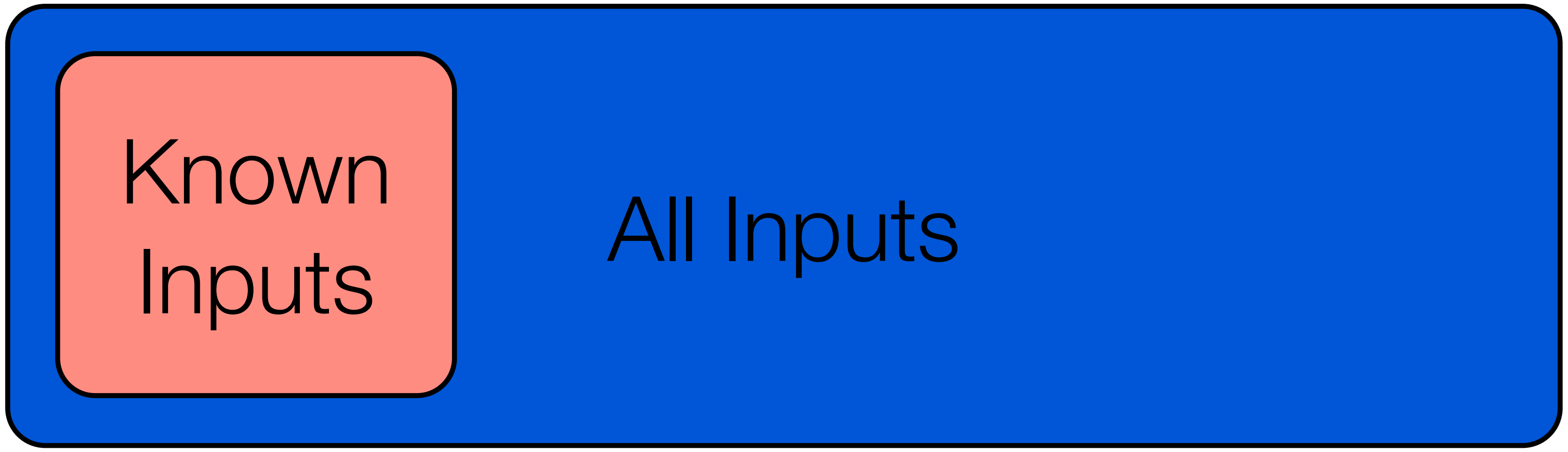
    - Memory assistance [Hodges et. al. 2006]

Image: CVPR 2011, Aghazadeh et. al.,link

Introduction : Memory Selection

- We must decide what visual inputs to remember.
- How should this be done?
    - Novelty detection.
- What is novelty detection?

Introduction : Novelty Detection



- *Novelty* = All Inputs - Known Inputs
- *Novelty detection*: identification of inputs that differ from previously seen inputs.
- Novelty detection can help decide on what is worth remembering.

Introduction : Setup

- Heuristic: detect novelty as deviation from background.
- Context: collect video sequences from from daily commute to work.
- Equipment: 4cm camera + memory stick.

Image: CVPR 2011, Aghazadeh et. al.,link

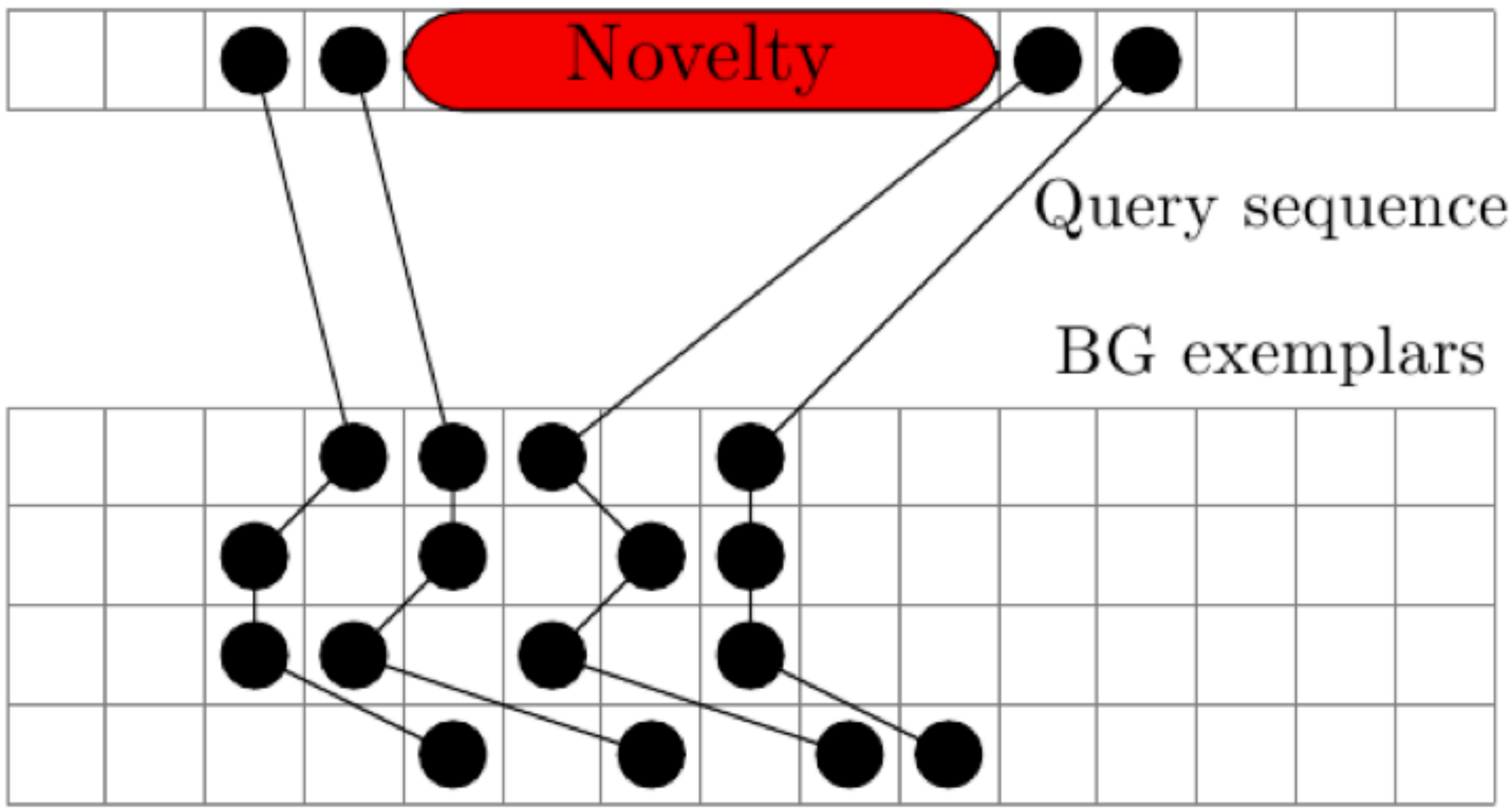
Introduction : Dataset

Dataset

- 31 videos of on average 5 minutes of a subject walking to work
- Each frame is manually labeled with a virtual location
- 4 sequences were manually identified to contain novelties
- Significant illumination/viewpoint variations
- Non-static environment

Image: CVPR 2011, Aghazadeh et. al.,link

Sequence Alignment



- Novelty is defined as a failure to register a sequence with a set of stored reference sequences (25 Hz videos sampled at 1 Hz.)

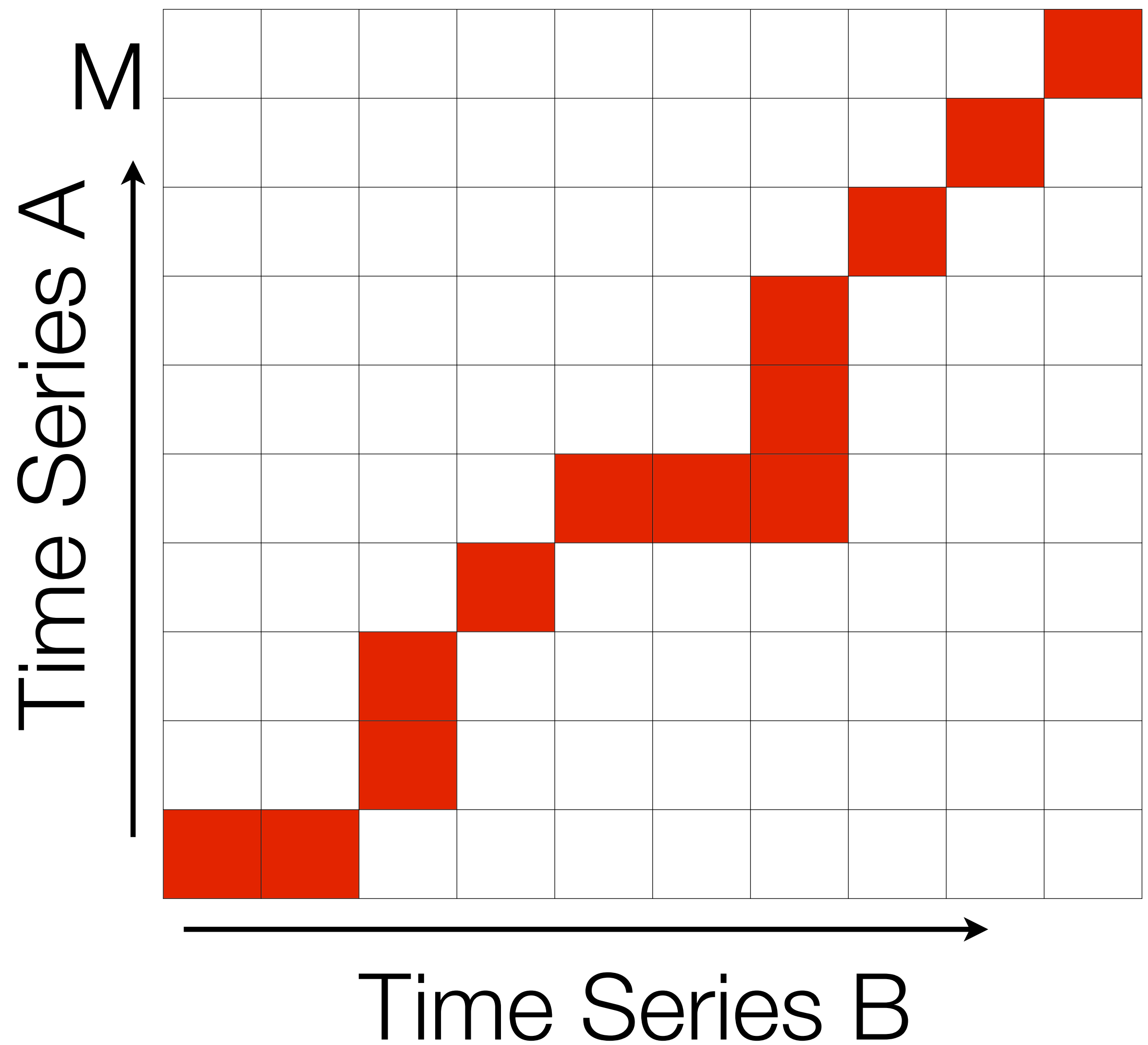
- Accomplished by sequence alignment, via Dynamic Time Warping (DTW).

Image: CVPR 2011, Aghazadeh et. al.,link

Sequence Alignment : Discussion

- Could we define or detect novelty in some other way?

Sequence Alignment : Dynamic Time Warping



Sequence Alignment : Similarity

- In order to use DTW, need to define some cost function
- This can by defining a measure of similarity between each pair of frames.
- Can use appearance based cues (SIFT, VLAD) to do this.

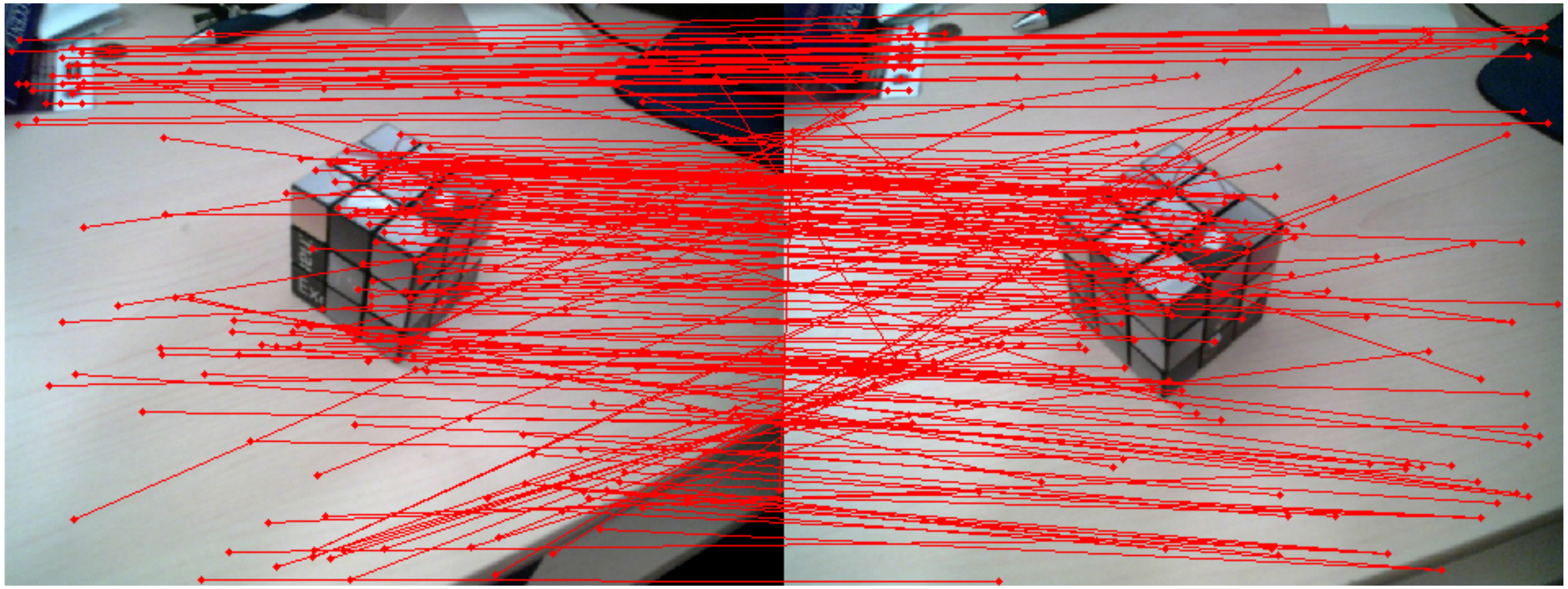


Image: link

Appearance Based Cues

- Can compute a fixed length vector each frame and use a kernel in order to compare similarity.
- Use SIFT or VLAD/SIFT to compute Bag of Features (BoF).
- VLAD: Vector of Locally Aggregated Descriptors:
  - (1) get k-means code book, and
  - (2) for each codeword C
    - take the L2-normalized sum of all the vectors assigned to it.

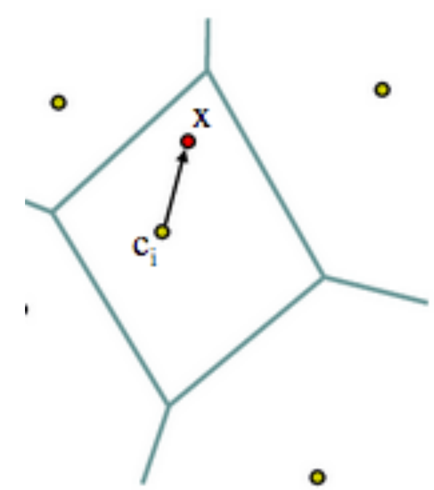


Image: link

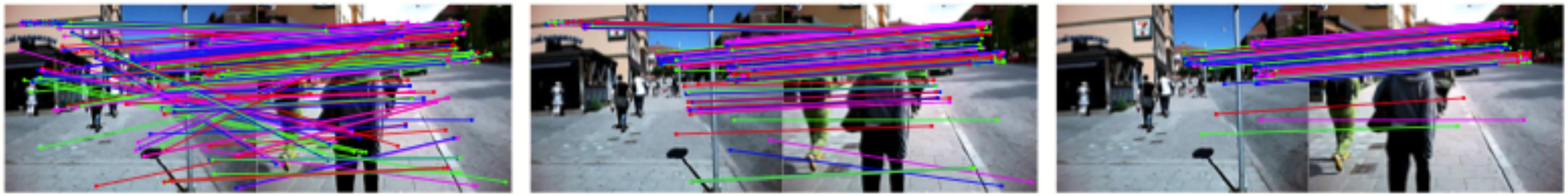
Geometric Similarity

- Appearance based cues alone are not accurate enough.
- Need to match local structures in a geometrically consistent way.
- Need a transformation that will do this: fundamental matrix.
- The measure of similarity will be the percentage of inliers in an initial set of putative matches, w.r.t to estimated fundamental matrix.
- Match against homography mapping to assess correctness of hypothetical fundamental matrix

Geometric Similarity : Discussion

- Could we supplement or substitute some other measure of similarity?
- How could different similarity measures affect novelty?

Example



250 putative matches | inliers wrt H | inliers wrt H and E

(meeting a friend)

Example

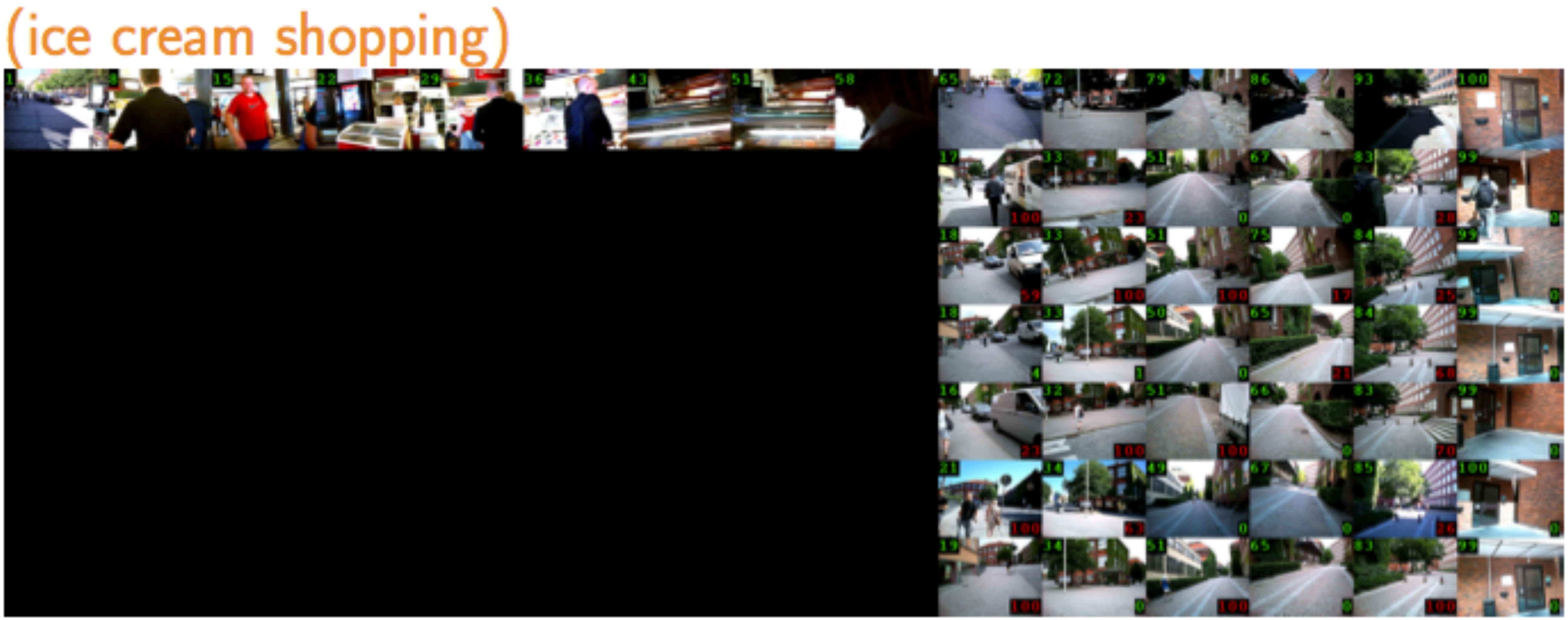
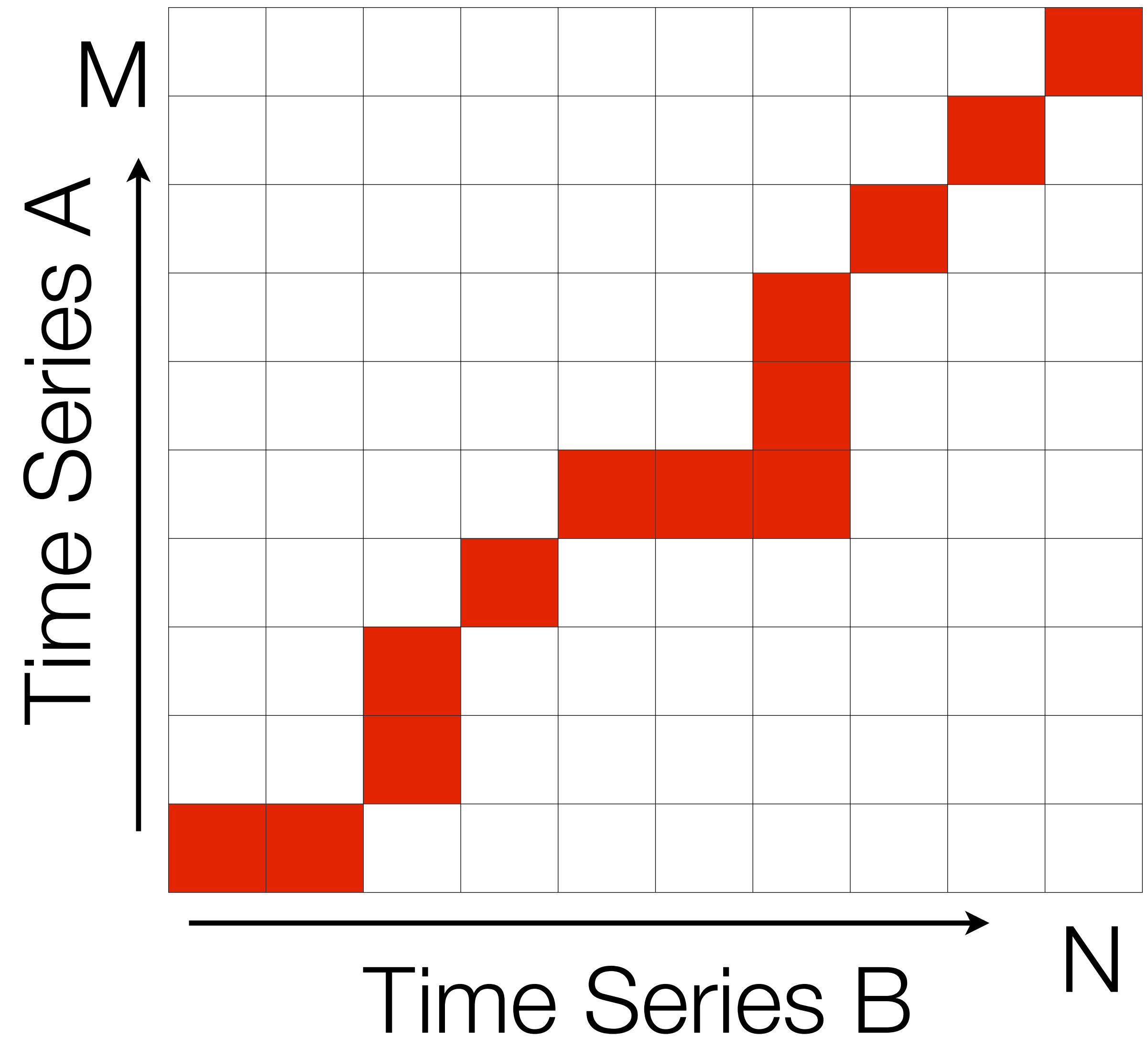


Image: CVPR 2011, Aghazadeh et. al.,link

Dynamic Time Warping

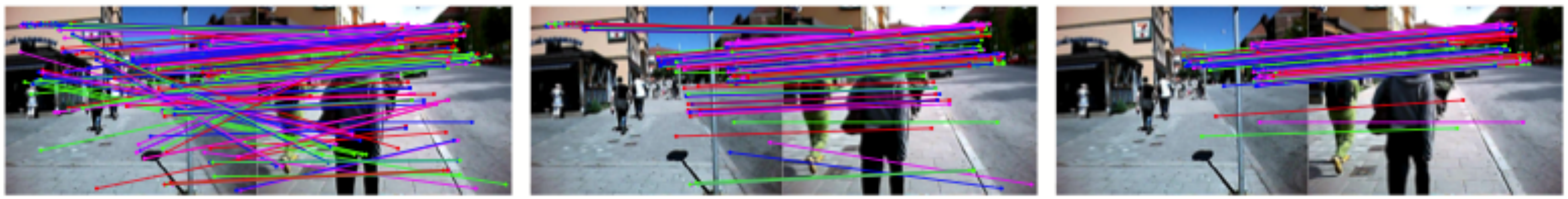
- Define a *path*:
  $p = \{(i_1, j_1), \ldots, (i_K, j_K)\}$
- s.t. (1) $(i_1, j_1) = (1, 1)$, and
  $(i_K, j_K) = (M, N)$

  (2) $p_{k+1} - p_k \in \{(0,1), (1,0), (1,1)\}$



- Define a cost function $c(i, j) \geq 0$
- Let $C_p = \sum_{k=1}^{K_p} c(i_k, j_k)$
- Want $p^* = \text{argmin}_p C_p$.
- Solved via dynamic programming.

Algorithm



250 putative matches | inliers wrt H | inliers wrt H and E

- compute features $\mathcal{F}_1, \mathcal{F}_2$ and nearest neighbor distance ratio
- keep best N matches P based on this ordering
- compute loose homography H_L and inliers P_H
- compute 5 point fundamental matrix E from P_H and inliers P_{HE}
- compute similarity $f_s = \min(1, \alpha \max(0, \frac{|P_{HR}|}{|P|} - \beta))$

Image: CVPR 2011, Aghazadeh et. al.,link

Algorithm : Cost Matrix

- Need to compute similarity matrix for sequences s_1 and s_2.
- Convert to cost matrix via zero-mean Gaussian with standard deviation σ_c.
  - Why? Noise?
- Use DTW to find optimal alignment!
- Problem: this is expensive.

Algorithm : Optimization

- Optimization: for each frame in s_1 find the k nearest neighbors in s_2 .
- Evaluate only the k nearest neighbors instead.

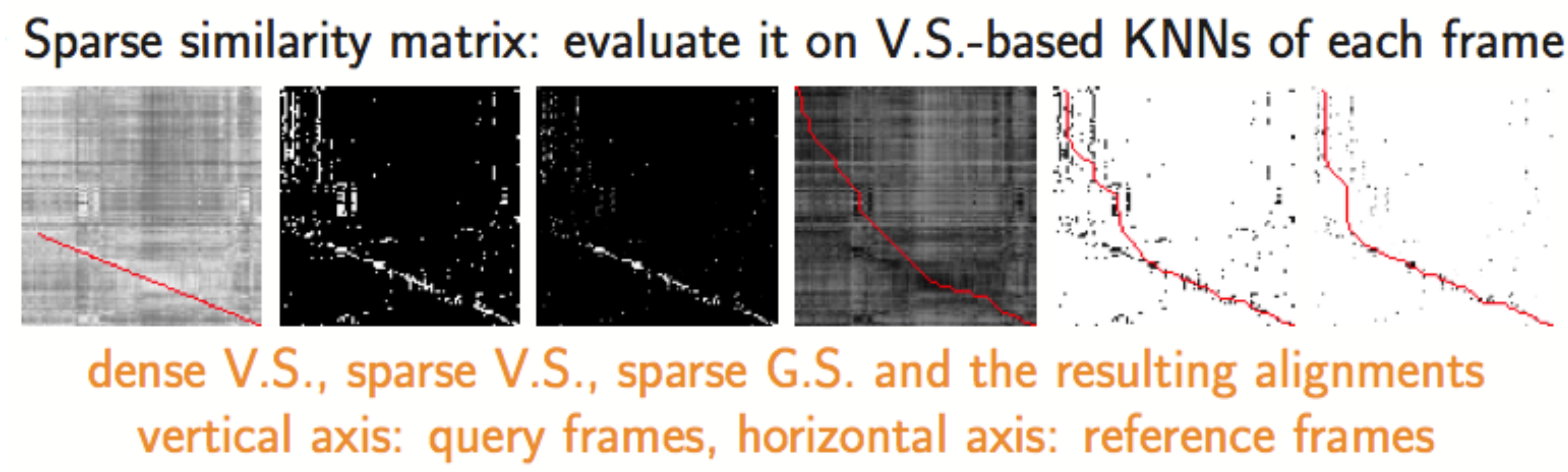


Image: CVPR 2011, Aghazadeh et. al.,link

Algorithm : Match Cost

- Let i correspond to frame indices in s_1 and j to frame indices in s_2.
- Let δ_{s_1,s_2} be the minimum cost path from DTW.
- The match cost $\lambda(i, \delta_{s_1,s_2})$ for a frame i in s_1 to s_2 is

$$\lambda(i, \delta_{s_1,s_2}) = \begin{cases} C_{i_k,j_k} & \text{if } \exists (i_k, j_k) \in \delta_{s_1,s_2} \text{ s.t. } i = i_k \\ 1 & \text{otherwise} \end{cases}$$

- where C_{i_k,j_k} is the value of the cost matrix at (i_k, j_k).

Algorithm : Novelty Detection

- Compute the minimum match cost for each frame in the query sequence:

$$E(s_t^{(i)}) = \min_{s_r \in S} \lambda(i, \delta_{s_q, s_r})$$

- where S contains all reference sequences.

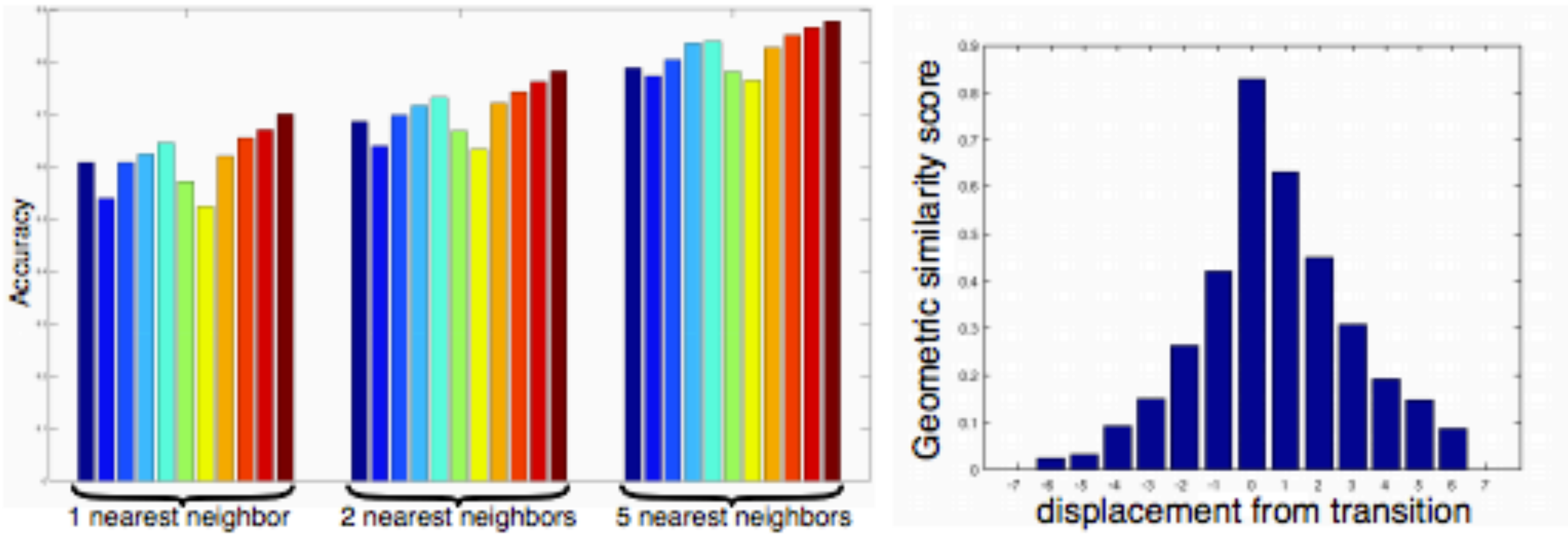
- Threshold the minimum match cost to find novelties.

- Smoothing: Gaussian mask applied to prior to matching with σ_N and using threshold $\Theta_N = e^{-\frac{1}{2^3 \sigma_c^2}}$.

Algorithm : Discussion

- How else could we implement memory selection or novelty detection?
  - How does this scale with the number of stored sequences?

Evaluation of Similarity Matching



(a) (b)

Figure 5: **(a)** The accuracy of image matching for differing interest region detectors and numbers of nearest neighbours. Methods (from left to right): VLAD+HessianAffine, VLAD+MSER, VLAD+HarrisAffine, VLAD+Dense(gray), VLAD+Dense(color), BoF+HessianAffine, BoF+MSER, BoF+HarrisAffine, BoF+Dense(gray), BoF+Dense(color), VLAD+BoF+Dense(gray+color). **(b)** The average of 100 F_{GV} values on local windows around the true correspondences.

- minimum intersection kernel for BoF and degree one polynomial kernel for VLAD/SIFT

- VLAD + BoF + Dense (gray + color) -> 88% = best

Image: CVPR 2011, Aghazadeh et. al.,link

Results : Detecting Novelty

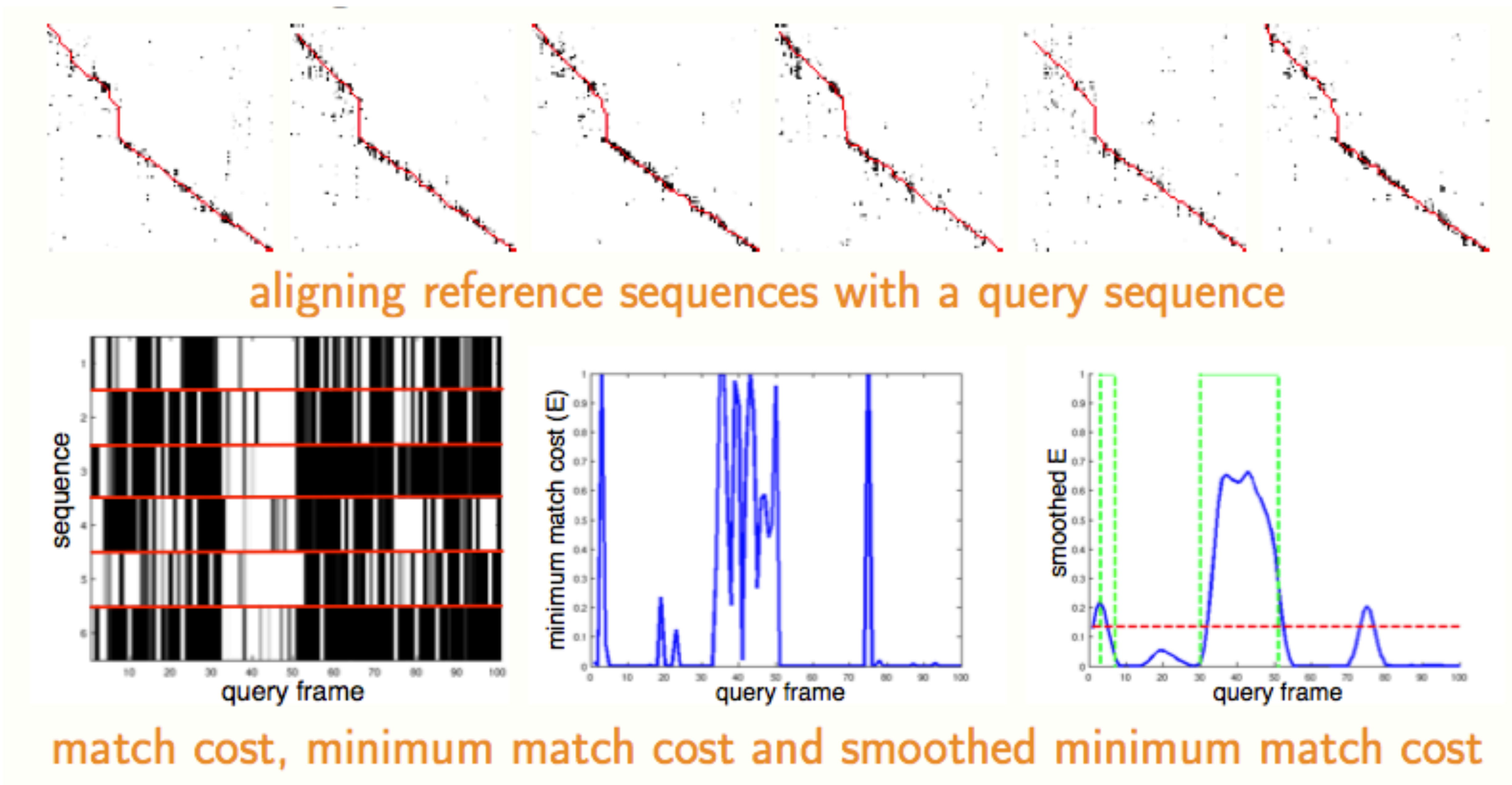


Image: CVPR 2011, Aghazadeh et. al.,link

Results : Precision Recall Curves and Matches

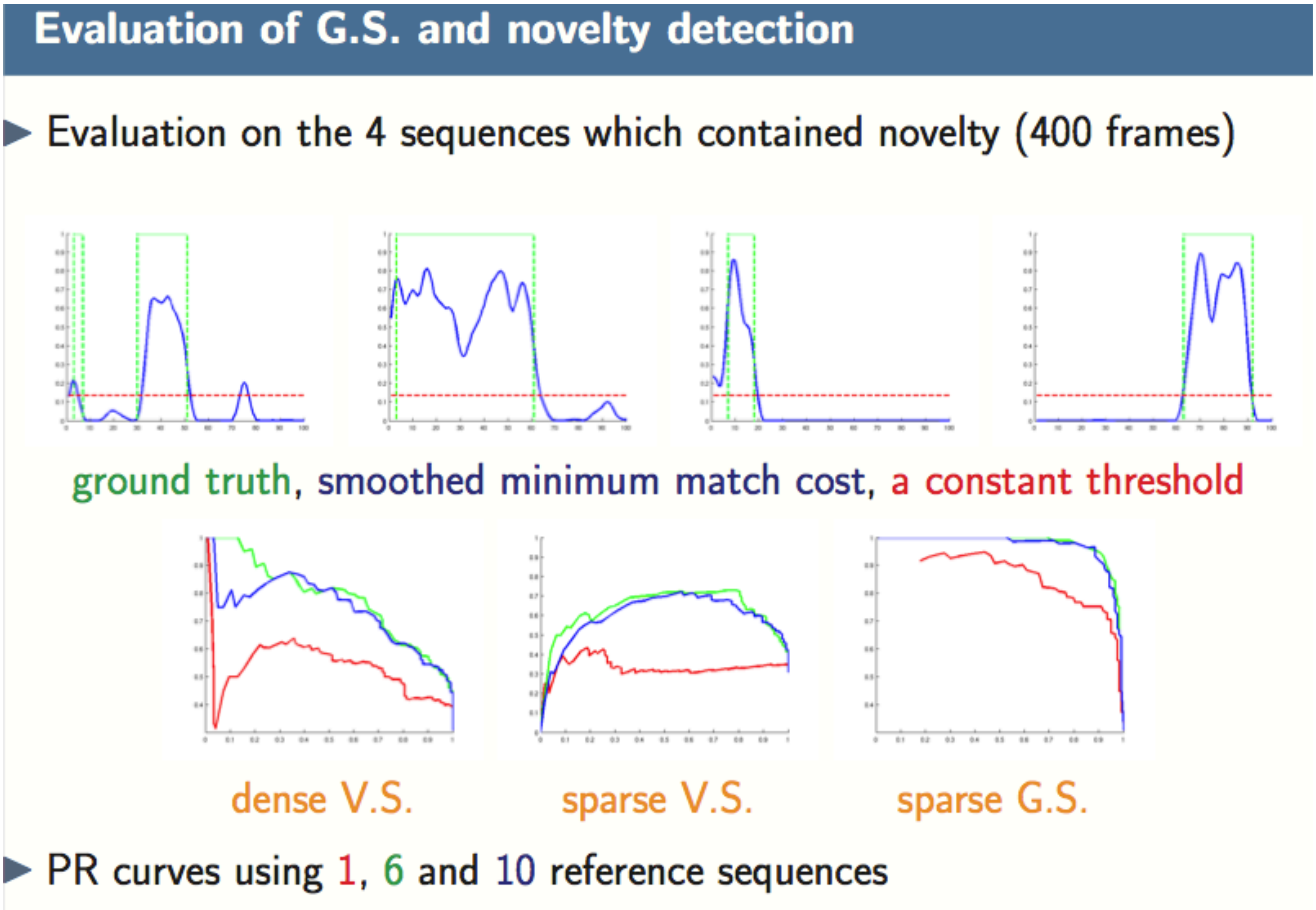


Image: CVPR 2011, Aghazadeh et. al.,link

Conclusion

- The scalability of this algorithm seems to be an issue.
- It would be interesting to explore alternative measures of similarity or novelty.
- Could this be converted to purely use clustering and only store clips for reference (by the user).
- The dataset is quite small, which is understandable given their technique, but perhaps an improved technique could make this work better?

References

- H. Jegou, M. Douze, C. Schmid, and P. Perez. Aggregating local descriptors into a compact image representation. In CVPR, 2010.
- M. Muller. Information retrieval for music and motion. Springer-Verlag New York Inc, 2007.
- Novelty Detection from an Egocentric Perspective. O. Aghazadeh, J. Sullivan, and S. Carlsson. CVPR 2011