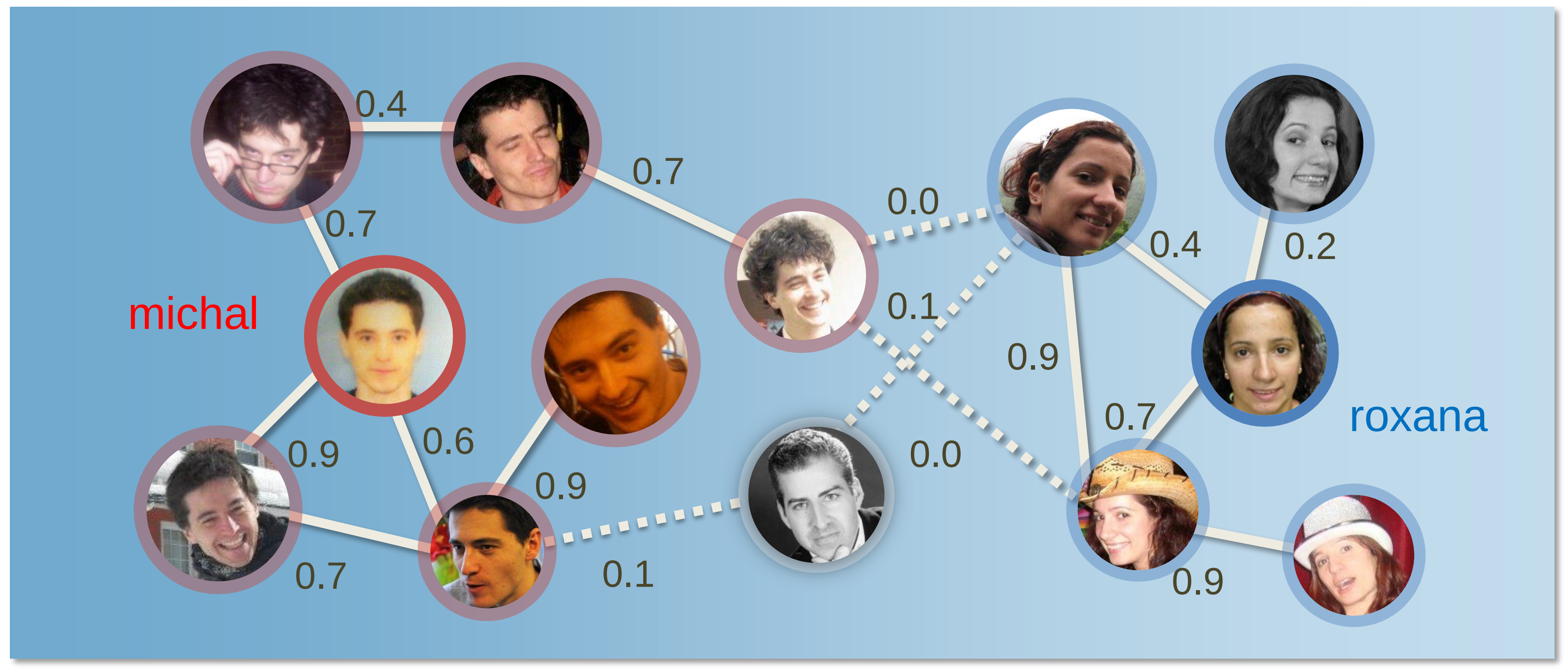
Robust Face Recognition Using Online Learning

MICHAL VALKO, BRANISLAV KVETON, LING HUANG, DANIEL TING, MATTHAI PHILIPOSE

University of Pittsburgh



An example of a similarity graph over faces. The faces are vertices of the graph. The edges of the graph connect similar faces. Labeled faces are outlined by solid lines.

Goal

A face recognition algorithm that:

- Has a high accuracy
- Has a high recall
- Is robust to outliers
- Runs in real time

Challenges

High accuracy and a high recall are contradicting objectives. Achieving both with standard ML algorithms is typically impossible (unless the training set closely resembles the test set).

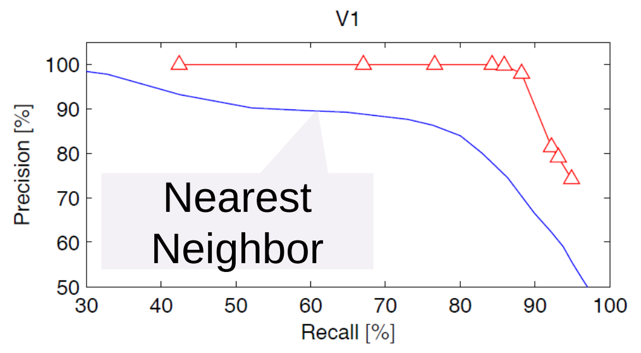
Adaptive ML algorithms can help to achieve this objective. The problem is that no explicit feedback (labels) is provided in real time.

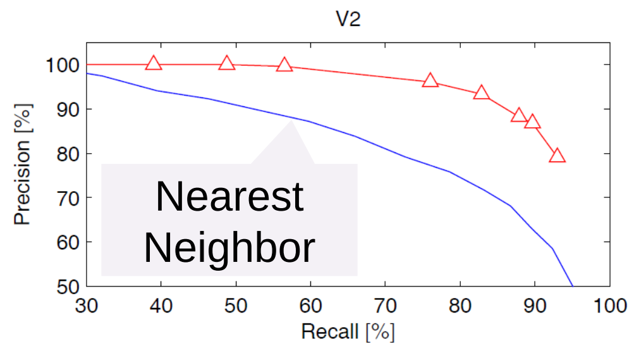
Approach

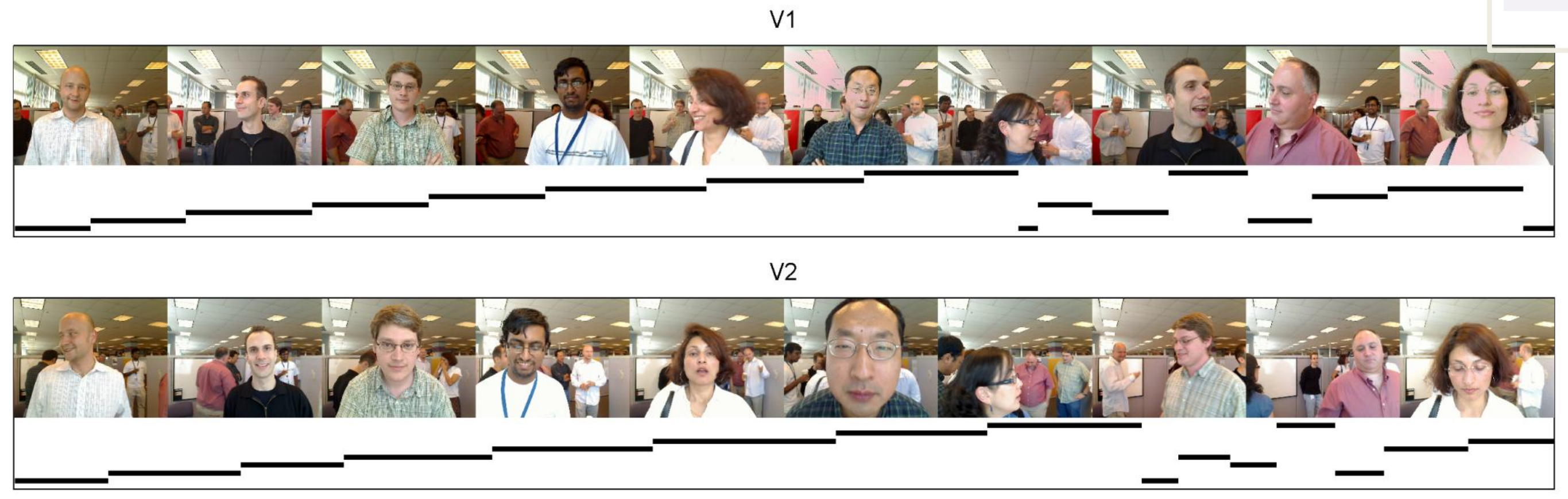
Online (real-time and incremental) learning of a similarity graph over observed faces and inference of face IDs based on the structure of the graph.

Algorithm (time step t):

- Remove outliers from the graph
- If the graph is too large, coarsen the adjacency graph by replacing the neighboring points with a set of local representative points
- Add the new face to the graph
- Infer the ID of the face based on the structure of the graph - a random walk that starts at the new face and terminates at the labeled faces







The horizontal lines denote time at which different people appear on the videos.

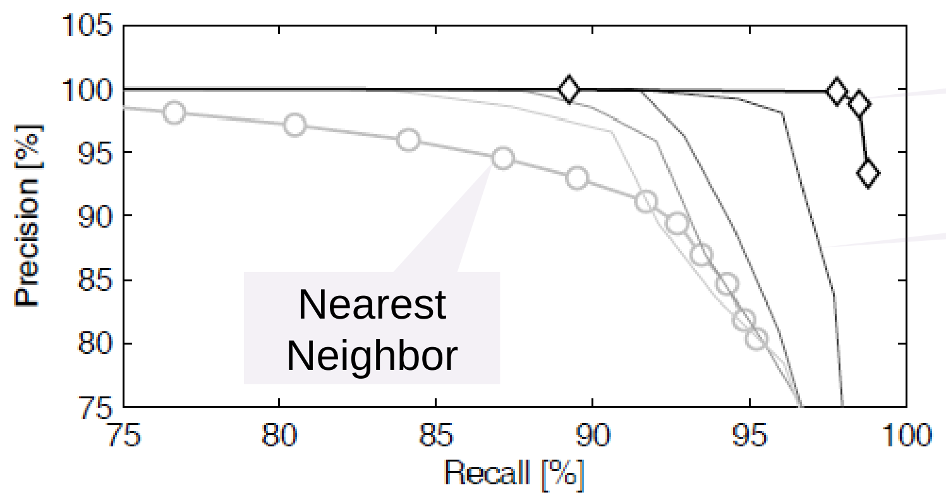
Online Algorithm

```
Inputs:
  an unlabeled example $\mathbf{x}_t$
  up to $n_g$ representative vertices $C_{t-1} = \{\mathbf{c}_1, \mathbf{c}_2, \dots\}$
  vertex multiplicities $\mathbf{v}_{t-1}$

Algorithm:
  $C_t = C_{t-1}$
  $\mathbf{v}_t = \mathbf{v}_{t-1}$
  while $(|C_t| = n_g + 1)$
    $R = 2R$
    greedily choose $C_t \subset C_{t-1}$ with $\min_{a,b \in C_t} \|a - b\| > R$
    update $\mathbf{v}_t$ based on the repartitioning
  if $\mathbf{x}_t$ is closer than $R$ to any $\mathbf{c}_i \in C_t$
    $\mathbf{v}_t(i) = \mathbf{v}_t(i) + 1$
  else
    $\mathbf{v}_t(|C_t| + 1) = 1$
    add $\mathbf{x}_t$ to the position $(|C_t| + 1)$ in $C_t$
  build a similarity matrix $W_t$ over the vertices $C_t$
  build a matrix $V_t$ whose diagonal elements are $\mathbf{v}_t$
  $\hat{W}_t = V_t W_t V_t$
  compute the Laplacian $\hat{L}$ of the graph $\hat{W}_t$
  infer labels on the graph:
    $\ell^{\mathrm{oq}}[t] = \arg\min_{\ell} \ell^{\mathsf{T}} (\hat{L} + \gamma_g V_t) \ell$
    s.t. $\ell_i = y_i$ for all labeled examples up to the time $t$
  make a prediction $\hat{y}_t = \operatorname{sgn}(\ell_t^{\mathrm{oq}}[t])$

Outputs:
  a prediction $\hat{y}_t$
  up to $n_g$ representative vertices $C_t = \{\mathbf{c}_1, \mathbf{c}_2, \dots\}$
  vertex multiplicities $\mathbf{v}_t$
```

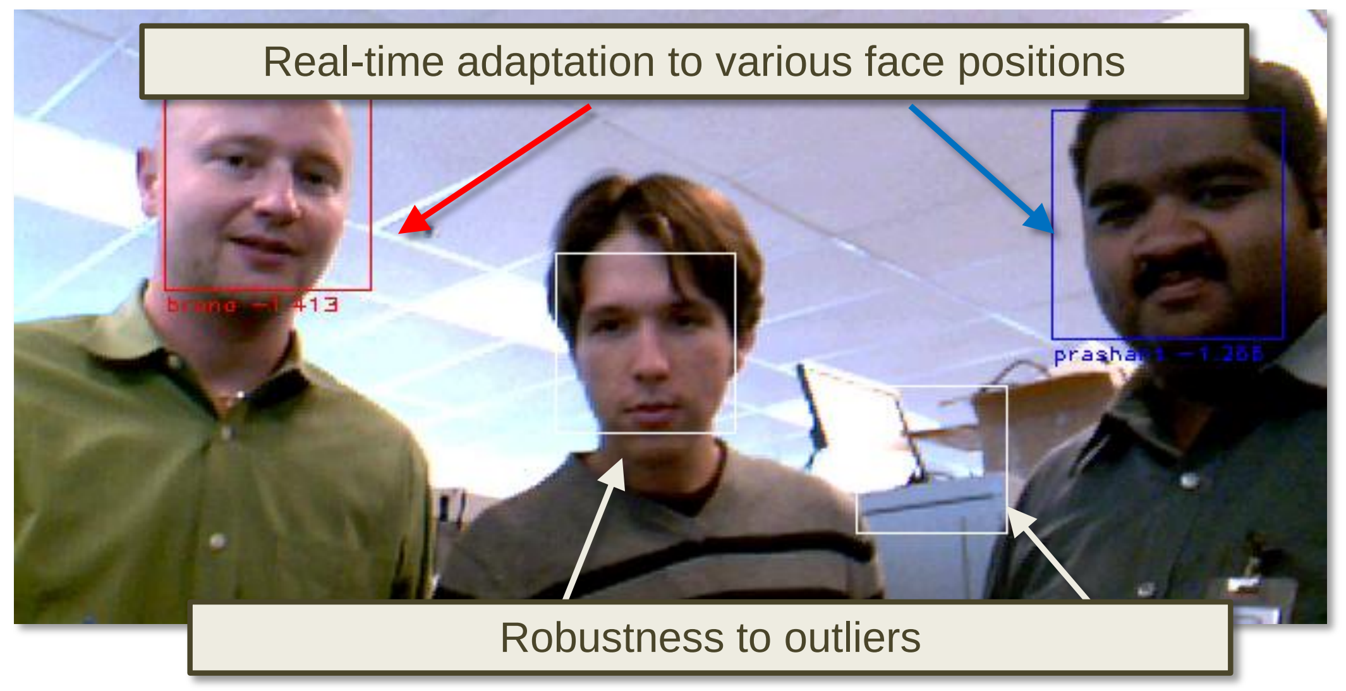
Online harmonic solution at the time step *t*. The main parameters of the algorithm is the regularizer γ_g and the maximum number of vertices n_g.



Our Algorithm

Online Semi-Supervised Boosting

Snapshots from the Adaptation dataset.
3 locations and 8 camera position.



Real-time adaptation to various face positions

Robustness to outliers

Similarity Matrix

- Defined over set of faces, higher weights to the pixels in the center
- $w_{ij} = \exp\left[-\frac{d^2(\mathbf{x}_i, \mathbf{x}_j)}{2\sigma^2}\right]$,

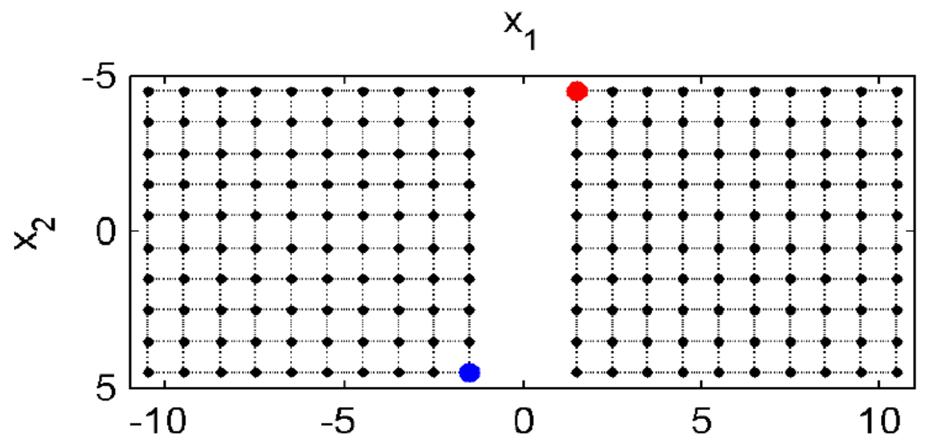
where $d(\mathbf{x}_i, \mathbf{x}_j) = \min \left\{ \begin{array}{l} \|\mathbf{x}_i - \mathbf{x}_j\|_{2,\psi}, \\ \|(\mathbf{x}_i - \bar{\mathbf{x}}_i) - (\mathbf{x}_j - \bar{\mathbf{x}}_j)\|_{2,\psi}, \\ \|\mathbf{x}_i / \bar{\mathbf{x}}_i - \mathbf{x}_j / \bar{\mathbf{x}}_j\|_{2,\psi} \end{array} \right\}$

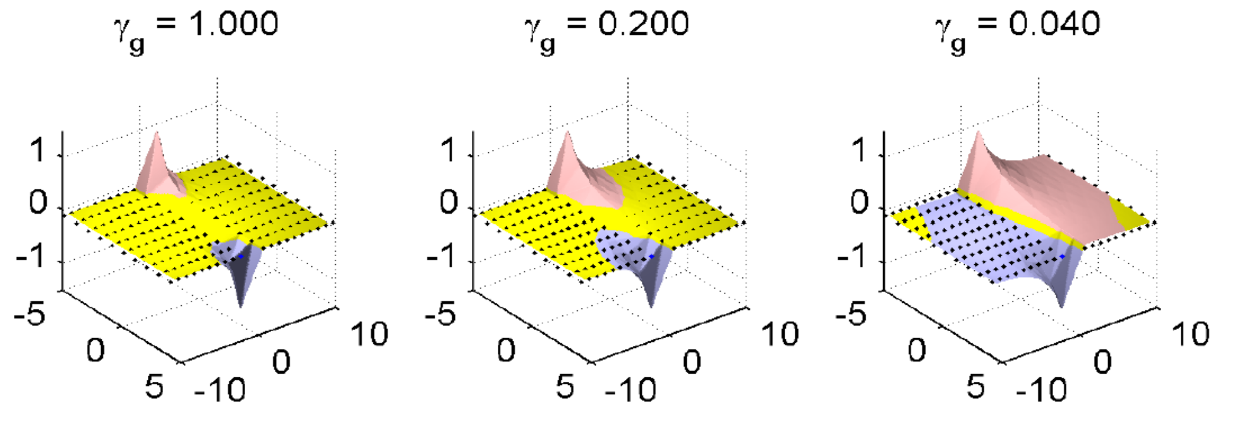
Data Quantization

- Cannot store all the past data
- Similarity graph needs to be reasonably small
- Use *k*-centers algorithm to maintain constant graph size
- Represent the multiple nodes by a single one
- Keep track of *multiplicities* → $\hat{\boldsymbol{\ell}}_u = (\hat{L}_{uu} + \gamma_g V)^{-1} \hat{W}_{ul} \boldsymbol{\ell}_l$

Regularized Harmonic Solution

Minimum satisfies the harmonic property and has a closed form solution.





Regularization controls the amount of extrapolation to unlabeled data. The smaller the regularizer, the more we trust unlabeled data.

Prediction Error Analysis

$$\frac{1}{n}\sum_{t=1}^{n}(\ell_t^{\mathrm{oq}}[t] - y_t)^2 \le \frac{9}{2n}\sum_{t=1}^{n}(\ell_t^{*} - y_t)^2 + \frac{9}{2n}\sum_{t=1}^{n}(\ell_t^{\mathrm{o}}[t] - \ell_t^{*})^2 + \frac{9}{2n}\sum_{t=1}^{n}(\ell_t^{\mathrm{oq}}[t] - \ell_t^{\mathrm{o}}[t])^2$$

True risk close to empirical by the algorithm stability argument of Cortes et al. (2008)

Difference between the offline and online prediction

Quality of quantization