Link Prediction in Email Exchange Network

David Brody, Roy Frostig, David Tobin

Department of Computer Science, Stanford University, Stanford, California

Link Prediction in Email Exchange Network

David Brody, Roy Frostig, David Tobin

Department of Computer Science, Stanford University, Stanford, California

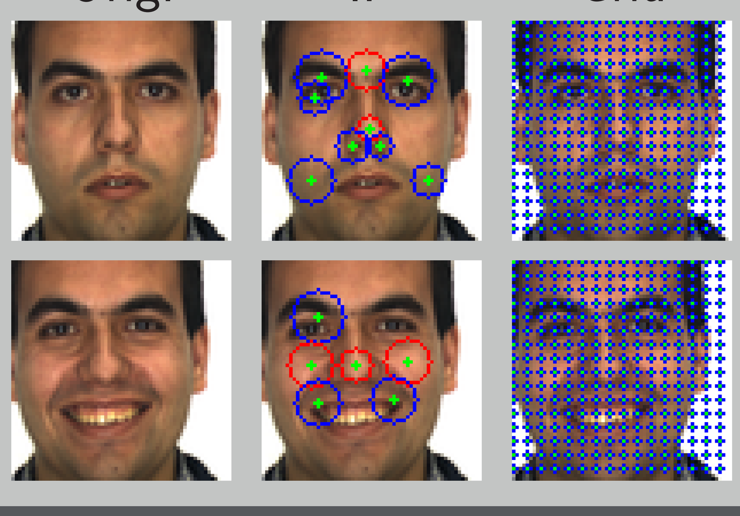
Introduction

We descibe a probablistic graphical model for predicting future communications in a social network given communications from the past.

- Current implementations are limited
  - Only take into account similarity of social profiles
  - Assign score to link as opposed to probability distribution
  - Difficult to extract general relationships within network
- We propose:
  - Feature extraction to classify nodes
  - Learn parameters for activity between classes
  - Allow for nontrivial influence between nodes and activity
  - Allow for inference of class type and new links

Feature Extraction

- Interest point based feature extraction
  - SIFT or SURF interest point detector
  - leads to a very sparse description
- Grid-based feature extraction
  - overlaid regular grid
  - leads to a dense description



Feature Description

- Scale Invariant Feature Transform (SIFT)
  - 128-dimensional descriptor, histogram of gradients, scale invariant
- Speeded Up Robust Features (SURF)
  - 64-dimensional descriptor, histogram of gradients, scale invariant
- face recognition: invariance w.r.t. rotation is often not necessary
  - rotation dependent upright-versions U-SIFT, U-SURF-64, U-SURF-128

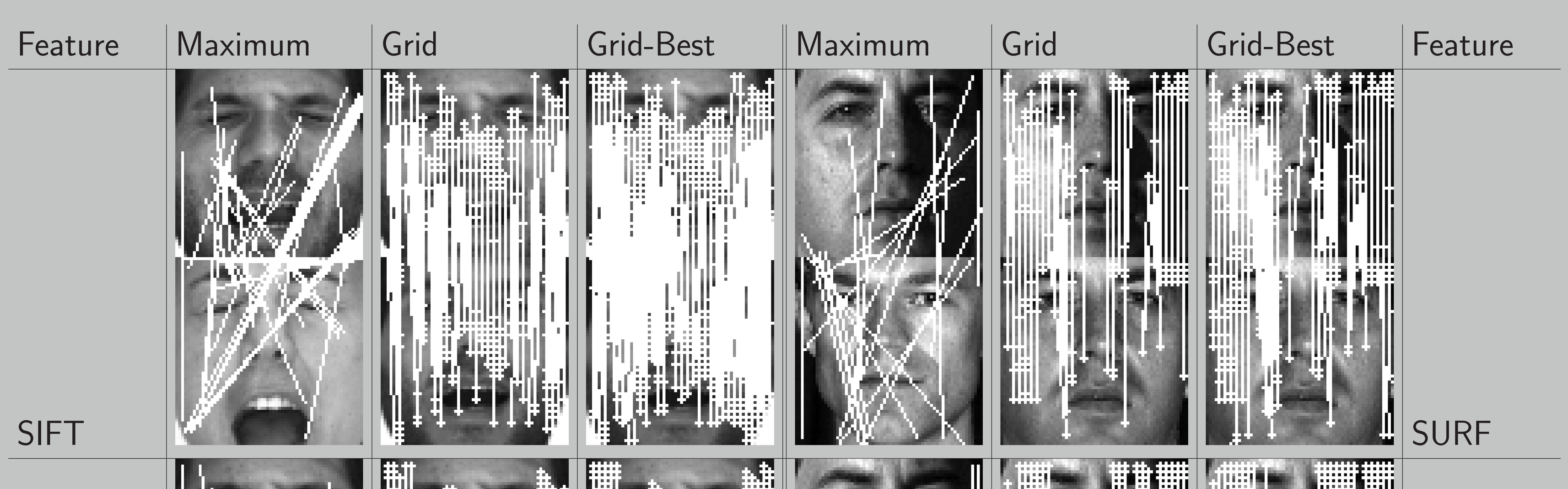
Feature Matching

- Recognition by Matching
  - nearest neighbor matching strategy
  - descriptor vectors extracted at keypoints in a test image $\mathbf{X}$ are compared to all descriptor vectors extracted at keypoints from the reference images $\mathbf{Y}_n, n = 1, \cdots, N$ by the Euclidean distance
  - decision rule:

$$\mathbf{X} \rightarrow r(\mathbf{X}) = \arg\max_c \left\{ \max_n \left\{ \sum_{x_i \in \mathbf{X}} \delta(x_i, Y_{n,c}) \right\} \right\}$$

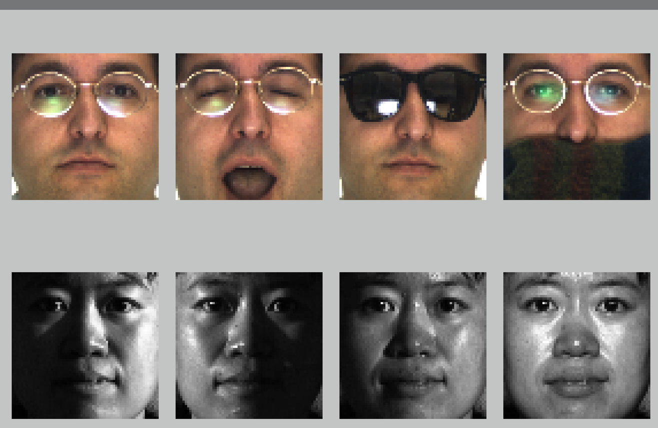
  - additionally, a ratio constraint is applied in $\delta(x_i, Y_{n,c})$
- Viewpoint Matching Constraints
  - maximum matching: unconstrained
  - grid-based matching: absolute box constraints
  - grid-based best matching: absolute box constraints, overlapping
- Postprocessing
  - RANSAC-based outlier removal
  - RANSAC-based system combination

Matching Examples for the AR-Face and CMU-PIE Database



Databases

- AR-Face
  - variations in illumination
  - many different facial expressions
- CMU-PIE
  - variations in illumination (frontal images from the illumination subset)



Results: Manually Aligned Faces

- AR-Face: 110 classes, 770 train, 770 test

Descriptor	Extraction	# Features	Error Rates [%]		
			Maximum	Grid	Grid-Best
SURF-64	IPs	64 × 5.6 (avg.)	80.64	84.15	84.15
SIFT	IPs	128 × 633.78 (avg.)	1.03	95.84	95.84
SURF-64	64x64-2 grid	64 × 1024	0.90	0.51	0.90
SURF-128	64x64-2 grid	128 × 1024	0.90	0.51	0.38
SIFT	64x64-2 grid	128 × 1024	11.03	0.90	0.64
U-SURF-64	64x64-2 grid	64 × 1024	0.90	1.03	0.64
U-SURF-128	64x64-2 grid	128 × 1024	1.55	1.29	1.03
U-SIFT	64x64-2 grid	128 × 1024	**0.25**	**0.25**	**0.25**

- CMU-PIE: 68 classes, 68 train (“one-shot” training), 1360 test

Descriptor	Extraction	# Features	Error Rates [%]		
			Maximum	Grid	Grid-Best
SURF-64	IPs	64 × 6.80 (avg.)	93.95	95.21	95.21
SIFT	IPs	128 × 723.17 (avg.)	43.47	99.33	99.33
SURF-64	64x64-2 grid	64 × 1024	13.41	4.12	7.82
SURF-128	64x64-2 grid	128 × 1024	12.45	3.68	3.24
SIFT	64x64-2 grid	128 × 1024	27.92	7.00	9.80
U-SURF-64	64x64-2 grid	64 × 1024	**3.83**	**0.51**	**0.66**
U-SURF-128	64x64-2 grid	128 × 1024	5.67	0.95	0.88
U-SIFT	64x64-2 grid	128 × 1024	16.28	1.40	6.41

Results: Unaligned Faces

- Automatically aligned by Viola & Jones

Descriptor	Error Rates [%]	
	AR-Face	CMU-PIE
SURF-64	5.97	15.32
SURF-128	5.71	11.42
SIFT	5.45	8.32
U-SURF-64	5.32	5.52
U-SURF-128	5.71	**4.86**
U-SIFT	**4.15**	8.99

- Manually aligned faces

- Unaligned faces

Results: Partially Occluded Faces

- AR-Face: 110 classes, 110 train (“one-shot” training), 550 test

Descriptor	Error Rates [%]					
	AR1scarf	*AR1sun*	*ARneutral*	*AR2scarf*	*AR2sun*	Avg.
SURF-64	2.72	30.00	0.00	4.54	47.27	16.90
SURF-128	1.81	23.63	0.00	3.63	40.90	13.99
SIFT	1.81	24.54	0.00	2.72	44.54	14.72
U-SURF-64	4.54	23.63	0.00	4.54	47.27	15.99
U-SURF-128	1.81	**20.00**	0.00	3.63	41.81	13.45
U-SIFT	**1.81**	20.90	**0.00**	**1.81**	**38.18**	**12.54**
U-SURF-128+R	1.81	19.09	0.00	3.63	43.63	13.63
U-SIFT+R	2.72	**14.54**	0.00	**0.90**	35.45	10.72
U-SURF-128+U-SIFT+R	**0.90**	16.36	**0.00**	2.72	**32.72**	**10.54**

Conclusions

- Grid-based local feature extraction instead of interest points