Fusing with Context: A Bayesian Approach to Combining Descriptive Attributes

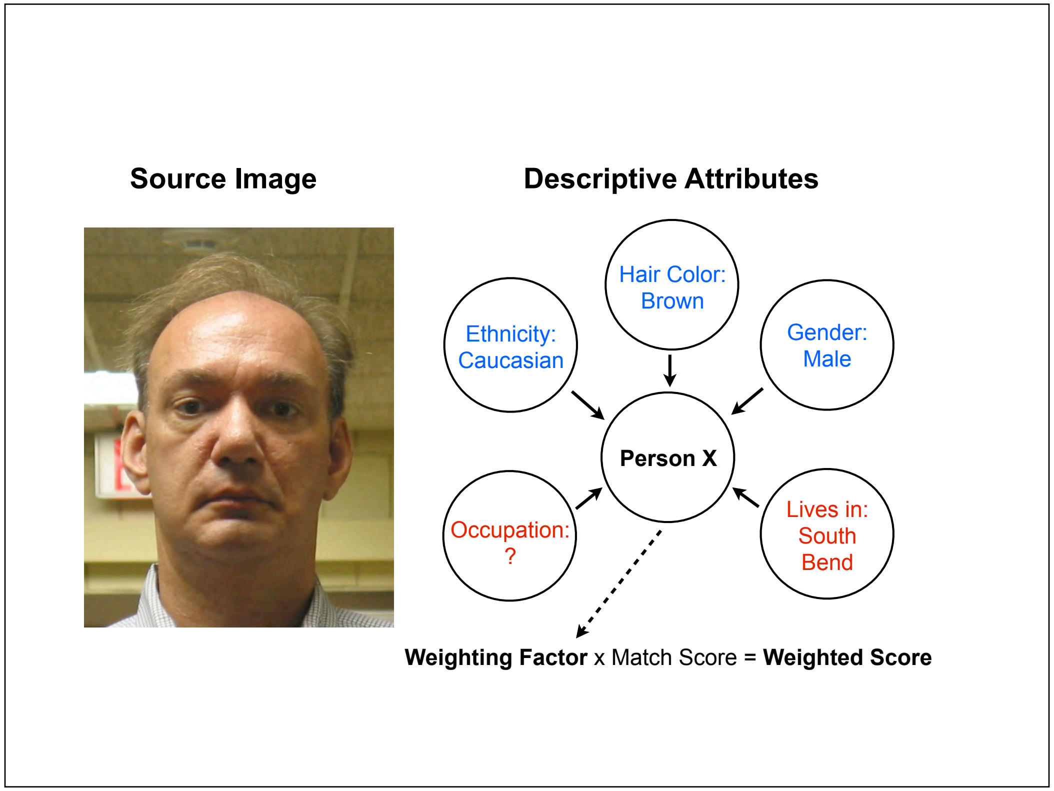
Walter J. Scheirer, Neeraj Kumar, Karl Ricanek,
Peter N. Belhumeur and Terrance E. Boult

This work was supported by ONR SBIR Award Number N00014-11-C-0243, and is an outgrowth of prior work supported by BIMA.

How well does unconstrained face identification work?

- Consider a top performing algorithm on the LFW set: V1-like features
  - Verification Rate[1]: 79.35%
  - Rank-1 Rate on a *subset*[2]: 41.64%
- What about very large populations?
  - UID controlled identification scenario for deduplication[3]: 600 Million enrollments calls for two modality fusion
    - Is this even enough? Time will tell...
- What kind of performance can we expect from a face algorithm operating in the real world?

1. http://vis-www.cs.umass.edu/lfw/results.html
2. A. Sapkota et al., "FACE-GRAB: Face Recognition with General Region Assigned to Binary Operator," IEEE Computer Society Workshop on Biometrics, 2010
3. R. Mashruwala and S. Prabhakar, "Multi-Modal Biometrics for One Billion People," Presentation at the IEEE Computer Society Workshop on Biometrics, 2010


Source Image
Descriptive Attributes
Hair Color: Brown
Ethnicity: Caucasian
Gender: Male
Person X
Occupation: ?
Lives in: South Bend
Weighting Factor x Match Score = **Weighted Score**

Contextual Attributes

- In many circumstances, contextual attributes are easy to obtain
- Consider a banking application designed to detect fraud
  - A user banks at Branch 1 95% of the time
  - The same user Banks in Region 1 100% of the time
  - Camera at ATM or teller's station captures face

Bank of America Logo (cc) by Neubie

Account is accessed at Branch 5 in Region 8 & recorded face does not match enrolled face: Fraud Detected

Image by Neubie "Bank of America Logo" BY http://www.davidneubert.com/

Prior Work: Visual Attributes

- Farhadi et al. CVPR 2009[1]
  - Describe objects by their *attributes*

Has Horn
Has Leg
Has Head
Has Wool

textual description

Mountain Goat (cc) by-nc-nd Cliff Hall

- Kumar et al. T-PAMI 2011[2]
  - Describe faces by their *attributes*

Has Hat
Has Beard
Has African Ethnicity
Has Round Nose

textual description

Ghostface Killah (cc) by-nc-nd Enrico Fuente

1. A. Farhadi, I. Endres, D. Hoiem, D. Forsyth, "Describing Objects by their Attributes," CVPR 2009
2. N. Kumar, A. Berg, P. Belhumeur, and S. Nayar, "Describable Visual Attributes for Face Verification and Image Search," IEEE T-PAMI, 2011

Top Image by Cliff Hall "Mountain Goat" BY-NC-ND http://www.flickr.com/photos/cliffhall/303337039/in/photostream/

Bottom Image by Enrico Fuente "Ghostface Killah" BY-NC-ND http://www.flickr.com/photos/okobojierik/5156583220/in/photostream/

Prior Work: Combining Disparate Information

- Bayesian Weightings and Networks
  - Well-known approach in the traditional security & intelligence domains
    - Wright et al. 2002[1]: military domain and doctrinal expertise
    - Laskey et al. 2004[2]: anomaly detection for document control
  - Bayesian weighting for biometrics
    - Jain et al. 2004[3]: "Soft Biometrics" + Traditional Biometric Matching
      - Information beyond soft biometrics?
      - Unknown variables?
      - Exponential growth of probability assignments?

1. E. Wright et al., "Multi-entity Bayesian Networks for Situation Assessment," 5th Intl. Conf. on Information Fusion, 2002
2. K. Laskey et al., "Detecting Threatening Behavior Using Bayesian Networks," Conf. on Behavioral Representation in Modeling and Simulation, 2004
3. A. Jain et al., "Soft Biometric Traits for Personal Recognition Systems," Intl. Conf. on Biometric Authentication, 2004

Bayesian Network Approach

Joint Probability Distribution:

$$p(\overset{\text{Network}}{\mathbf{x}}) = \prod_{i=1}^{N} p(\overset{\text{Variable}}{x_i}|\overset{\text{Parents}}{\mathrm{pa}_i})$$

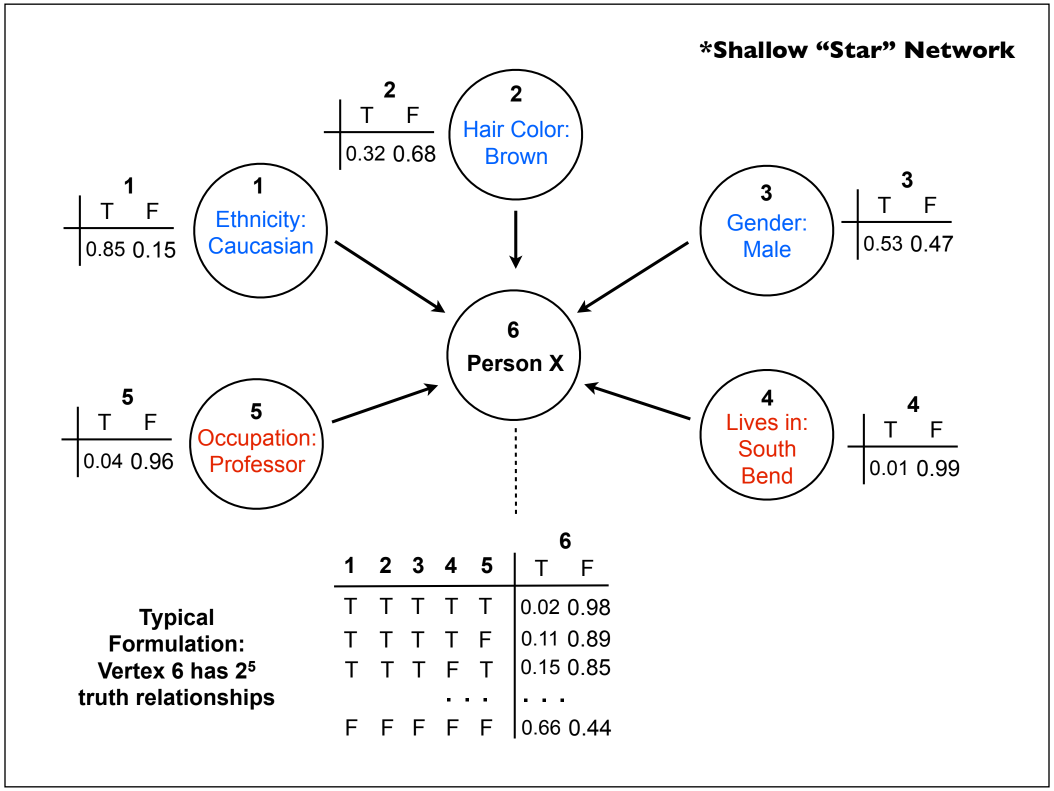
But we need to handle unknowns:

$$\underset{\text{User}}{p(\mathcal{U}}|\mathbf{x}_k) = \frac{\sum_{\mathbf{x}_u \in \{T,F\}} p(\overset{\text{Known Variables}}{\mathbf{x}_k} = T, \mathbf{x}_u, \mathcal{U} = T)}{\sum_{\mathbf{x}_u, \mathcal{U} \in \{T,F\}} p(\mathbf{x}_k = T, \underset{\text{Unknown Variables}}{\mathbf{x}_u}, \mathcal{U})}$$

Weighting:

$$\{(\overset{\text{Score from Biometric Alg.}}{s_1} \times \overset{\text{Probability from Bayesian Net}}{w_1}), (s_2 \times w_2), \ldots, (s_n \times w_n)\}$$

Note: There is a mistake in Eq. 2 of the paper; it should take the form of the middle Eq. in this slide.


*Shallow "Star" Network
1 Ethnicity: Caucasian
1 T F 0.85 0.15
2 Hair Color: Brown
2 T F 0.32 0.68
3 Gender: Male
3 T F 0.53 0.47
4 Lives in: South Bend
4 T F 0.01 0.99
5 Occupation: Professor
5 T F 0.04 0.96
6 Person X
Typical Formulation: Vertex 6 has 2^5 truth relationships
1 2 3 4 5 | 6 T F
T T T T T | 0.02 0.98
T T T T F | 0.11 0.89
T T T F T | 0.15 0.85
. . . | . . .
F F F F F | 0.66 0.44

Noisy-OR Approximation

- Two Problems with the Traditional Formulation:
  - It is not practical to assign all of the truth values by hand
    - $2^5 = 32$ assignments; $2^6 = 64$ assignments...
  - For many descriptive attributes, the causal interaction model is illogical
    - Example: Brown hair is independent of employment as a professor

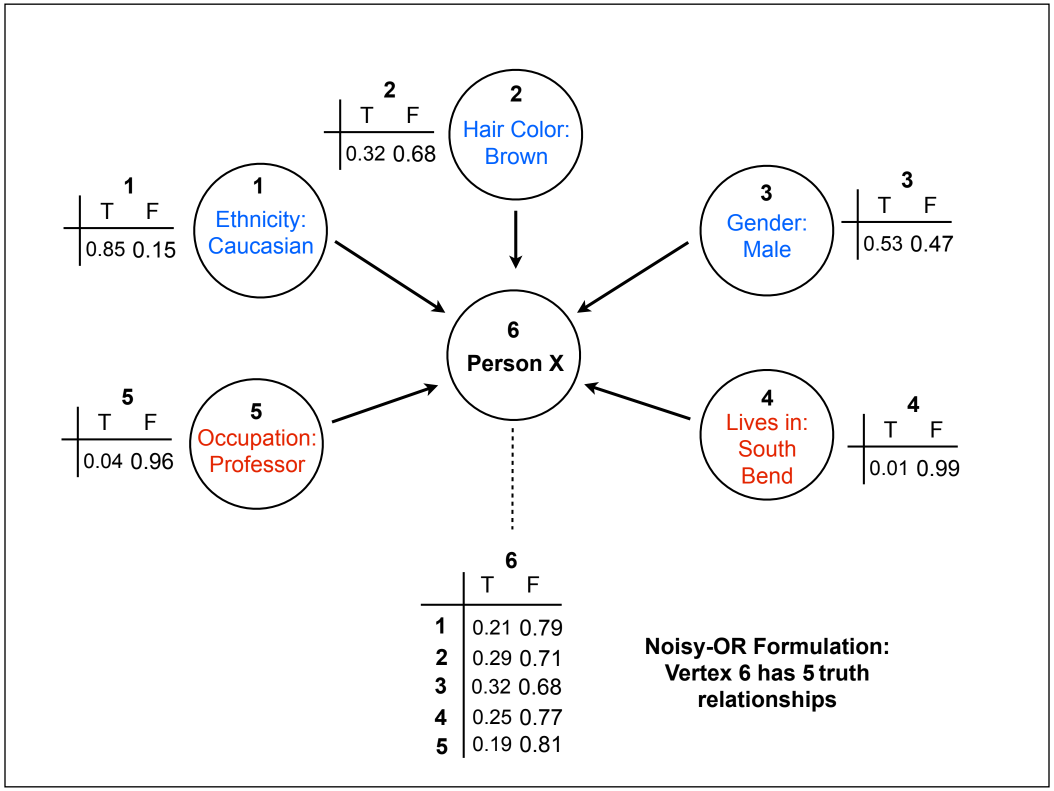
- Solution: decouple attributes from one another

Product of Probabilities:

$$p(\mathcal{U}|\mathbf{x}_k) = 1 - \prod_{i=1}^{n}(1 - p_i)$$

where

$$p_i = p(\mathcal{U} = T | x_i = T, \{\underset{\substack{\uparrow \\ \text{Known or Unknown}}}{x_j} = F\}_{j=1, j \neq i}^{n})$$



1
T F
0.85 0.15
1
Ethnicity: Caucasian
2
T F
0.32 0.68
2
Hair Color: Brown
3
Gender: Male
3
T F
0.53 0.47
6
Person X
5
T F
0.04 0.96
5
Occupation: Professor
4
Lives in: South Bend
4
T F
0.01 0.99
6
T F
1 0.21 0.79
2 0.29 0.71
3 0.32 0.68
4 0.25 0.77
5 0.19 0.81
Noisy-OR Formulation: Vertex 6 has 5 truth relationships

Robust Age Estimation

- Chen et al. AFGR 2010[1]
  - Statistical modeling of detected fiducial points
    - Facial features: shape, weight, sagging
    - Texture features: wrinkles, lines, ptosis
    - Perform Support Vector Regression over extracted features
  - Training Data: 600 adult faces from MORPH[2] and PAL[3]

Selection of Images from MORPH, K. Ricanek and T. Tesafaye, 2006

1. C. Chen et al., "Face Age Estimation Using Model Selection," IEEE AMFG, 2010
2. K. Ricanek and T. Tesafaye, "MORPH: A Longitudinal Image Database of Normal Adult Age-Progression," IEEE AFGR, 2006
3. M. Minear and D. Park, "A Lifespan Database of Adult Facial Stimuli," *Behavior Research Methods, Instruments & Computers*, 36(4), 2004

Visual Facial Attributes

- Kumar et al. 2011
  - Low-level simple features + machine learning
    - Feature extractors are composed of pixels from face region, pixel feature type, normalization and aggregation
    - From an aligned image I, extract low level features:

$$\mathcal{F}(I) = \{\mathbf{f}_1(I), \ldots, \mathbf{f}_k(I)\}$$

    - In total, we trained **73** different SVM attributes classifiers
    - Crowdsourced ground truth labeling; 500-2000 +/- examples from the Columbia Face Database

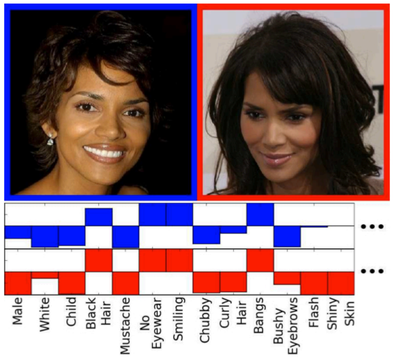


Image adapted from Fig 1. in N. Kumar et al. "Describable Visual Attributes for Face Verification and Image Search," *T-PAMI*, 2011

Experiments

- Two Data Sets
  - MBGC: Faces + Simulated Context
  - Celeb DB: Faces + Real Context
- Two Face Algorithms
  - LRPCA[1]
  - V1-like Features[2]
- Metric: Percentage Improvement

$$\%\mathcal{I}_n = (\text{Weighted } \underset{\substack{\uparrow \\ \text{Recognition Rate}}}{\mathcal{R}_n} - \text{Original } \mathcal{R}_n)/\text{Original } \mathcal{R}_n$$

1. http://www.cs.colostate.edu/facerec/algorithms/lrpca2010.php
2. N. Pinto et al., "How Far Can You Get on a Modern Face Recognition Test Using Only Simple Features?" IEEE CVPR, 2011

Experiment 1: Attributes & Accuracies

Visual Attrs. & Accuracy	Contextual Attrs.
Age (+/- 7 years); 89.9%	Lives in city X
Gender; 86.7%	Works as X
Eyeglasses; 96.6%	Works at X
Weight: Chubby; 87.8%	Has n children
Eyebrows: Bushy; 88.2%	Is the mother of X
Hair Color: Black; 92.3%	Is the brother of X
Hair Color: Brown; 86.5%	Frequents bank X
Ethnicity: Asian; 94.6%	Owns a car
Ethnicity: African; 97.4%	Attends school X
Ethnicity: European; 87.1%	Graduated in X

Gender: Male
Ethnicity: Asian
Hair: Black
Wearing Eyeglasses
Estimated Age: 28

Gender: Female
Ethnicity: European
Hair: Brown
Not Wearing Eyeglasses
Estimated Age: 22

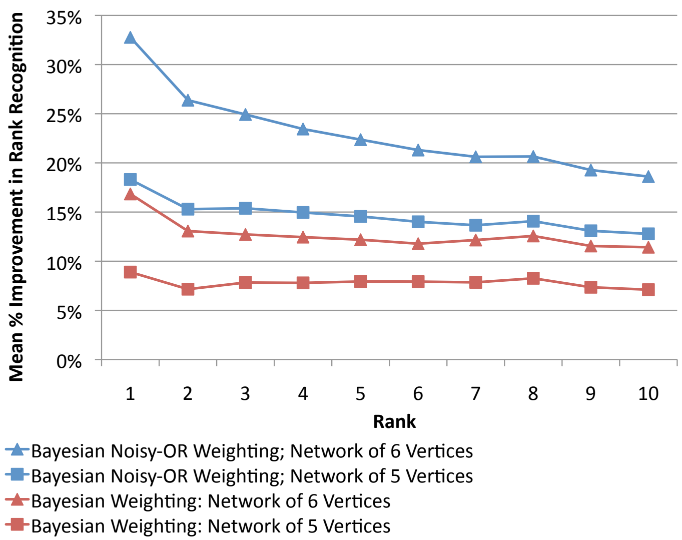
Gender: Male
Ethnicity: African
Eyebrows: Bushy
Weight: Skinny
Estimated Age: 34

Experiment 1: Simulated Context

- LRPCA & MBGC*: 466 unique people; 217,156 scores
- Two network sizes: 5 & 6 vertices
- 10 different combinations of attributes were chosen for each experiment
- Each attribute is used at least once
- Consistency between the typical and Noisy-OR formulations
- All visual attributes are generated from the source probe and enrollment images
- The set of observations from the probe is left incomplete: 1 variable from the enrollment network is always unknown

*P. Phillips et al., "Overview of the Multiple Biometrics Grand Challenge," ICB, 2009

Experiment 1: Simulated Context



Experiment 1: Simulated Context

Summary of mean rank-1 accuracies ($\overline{\mathcal{R}}_1$) and mean percentage improvement ($\%\overline{I}_1$) for LRPCA weighting

Experiment	**Vertex Count**	**CPT Entries**	$\overline{\mathcal{R}}_1$	$\%\overline{I}_1$
Baseline LRPCA	-	-	65.9%	-
Bayesian Weighting	5	20	71.7%	8.9%
Bayesian Weighting	6	37	77.0%	18.8%
Bayesian Noisy-OR Weighting	5	9	77.9%	18.3%
Bayesian Noisy-OR Weighting	6	10	87.5%	32.8%

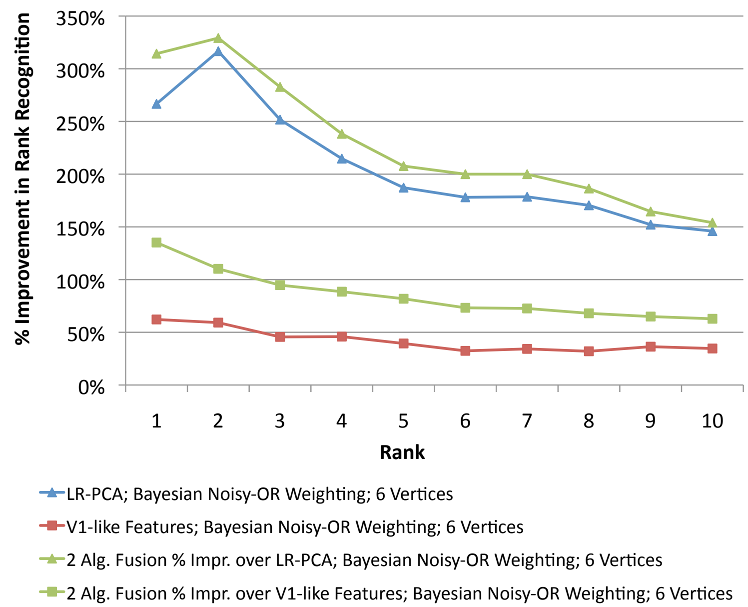
Experiment 2: Real Context

- Celeb DB[1]: Face Image + Biographies of Celebrities
  - We chose 167 unique people
- LRPCA, V1-like features, and w-score[2] fusion of both algorithms
- Attributes: Eyeglasses, Gender, Bushy Eyebrows, "Known Associates", First Film
- All visual attributes are generated from the source probe and enrollment images
- The set of observations from the probe is left incomplete: 1 variable from the enrollment network is always unknown

Known Associates: James Cagney, Lauren Bacall, Edward G. Robinson Bette Davis, Peter Lorre

1. Bolme et al., "Person Identification Using Text and Image Data," IEEE BTAS, 2007
2. Scheirer et al., "Robust Fusion: Extreme Value Theory for Recognition Score Normalization," ECCV, 2010

Experiment 2: Real Context



Conclusions

- Tools for effective identity centric Bayesian Nets
  - Non-image context
  - Descriptive visual attributes
  - Shallow "star" network
  - Noisy-OR approximation

- Areas for future work
  - Enhance and expand visual attributes
  - Multiplicative score weighting is likely not the best strategy
    - Apply weight at the sensor or feature level
    - Combine with a probabilistic normalization approach (w-score)

Questions?