

Information content of biometric features

Andy Adler, Richard Youmaran, Sergey Loyka

University of Ottawa,
Canada

Question:

How much information is in a biometric measurement?

Or

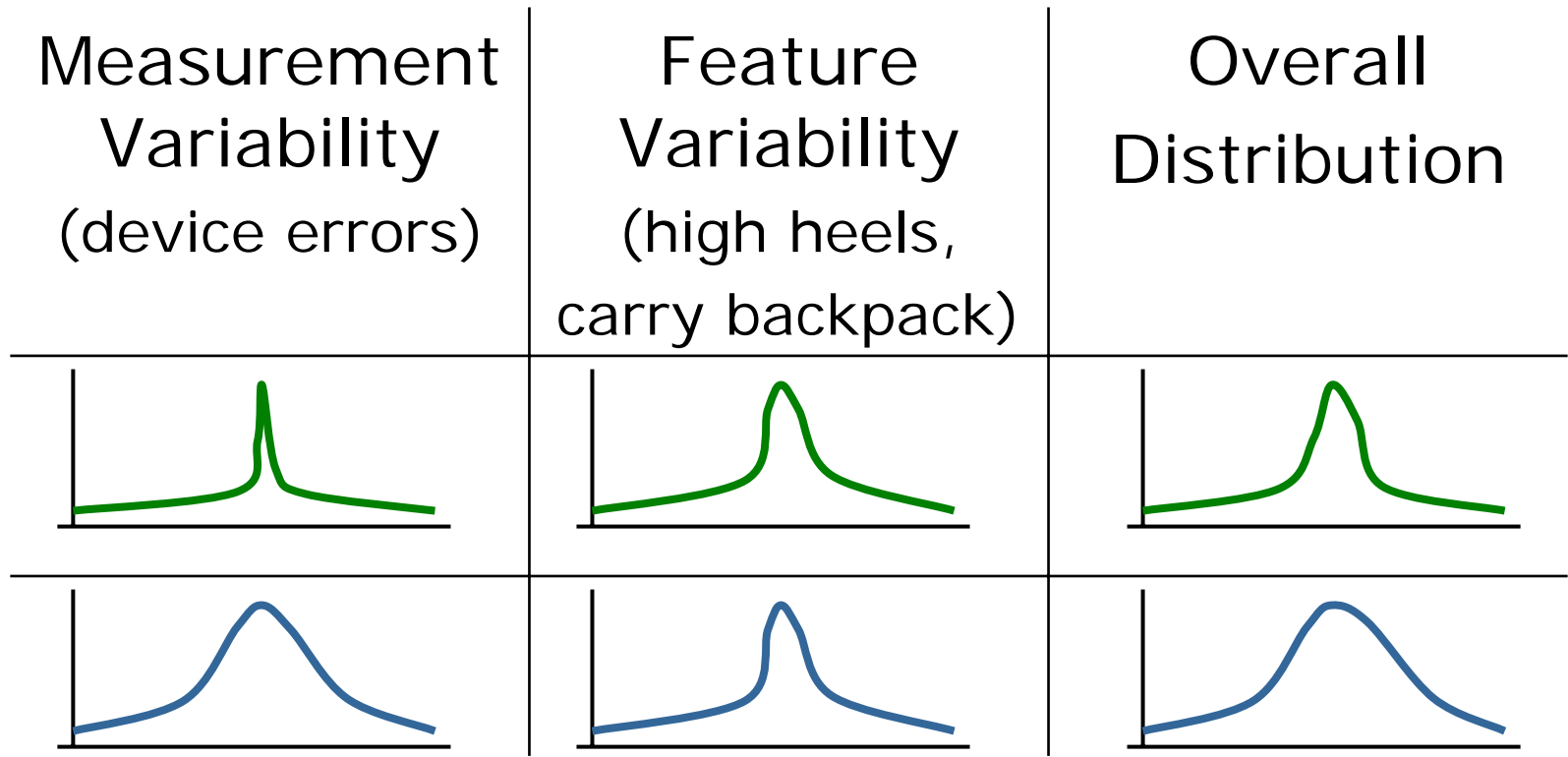
How much do we learn (about identity) from a biometric image

Or

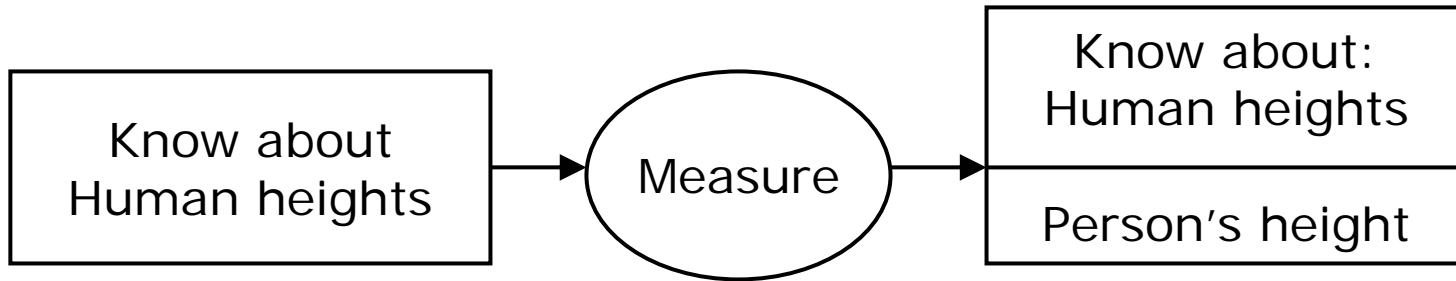
How much privacy do we lose on releasing a biometric image

Example: measure *Height*

- **Measure #1** (at doctor's office, ie. accurate)
- **Measure #2** (via telescope, ie. inaccurate)



Example: measure *Height*



□ How much information learned?

	Average (5½' tall)	Tall (7½' tall)
Measure #1	Low	Quite a lot
Measure #2	Almost zero	Low

Proposed measure:

relative entropy $D(p||q)$

- Given biometric feature vector \mathbf{x}
- Distributions
 - intra-person distribution, $p(\mathbf{x})$
 - inter-person distribution, $q(\mathbf{x})$
- $D(p||q)$ measures inefficiency of assuming q when true distribution is p

Or,

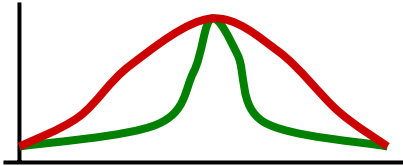
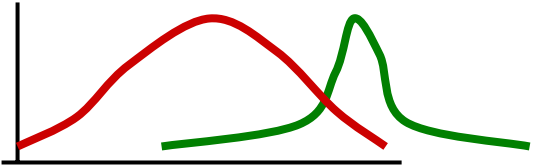
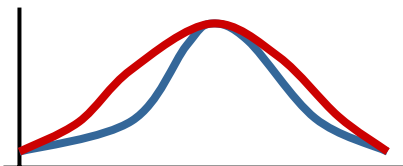
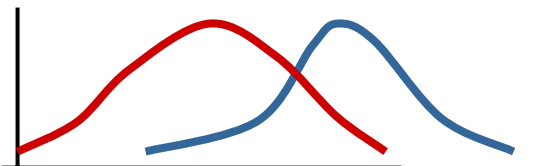
- $D(p||q)$ measures extra information in p than q

Relative entropy

$$D(p \parallel q) = \int p(\mathbf{x}) \log_2(p(\mathbf{x}) / q(\mathbf{x})) d\mathbf{x}$$

- Need to calculate feature distributions
 - Population (q)
 - Individuals (p)
- Real features are correlated (eg. height and weight)
- Entropy, $H(\mathbf{x})$, is not appropriate
 - Doesn't account for individual variability (p)
 - Example: if hair is blue in face image, that is not information for identity

Relative Entropy for height

	Average (5½' tall)	Tall (7½' tall)
Measure #1 (accurate)	 <p>$D = 0.23$ bits</p>	 <p>$D = 2.7$ bits</p>
Measure #2 (inaccurate)	 <p>$D = 0.05$ bits</p>	 <p>$D = 1.1$ bits</p>

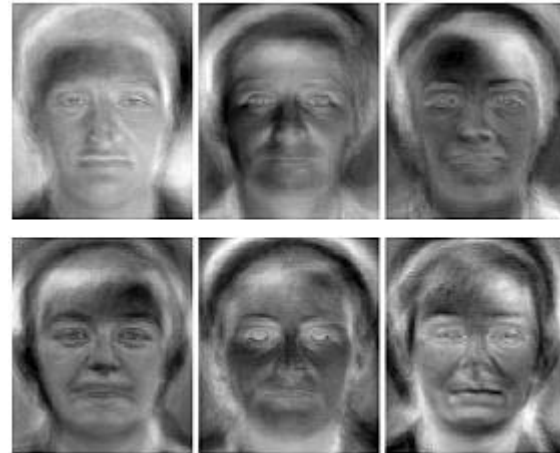
Example 2: *Face Recognition*

Motivation:

- Eigenface feature dimensionality is constant (unlike FP minutiae)
- Our work with biometric encryption with FR fit about 20 bits of key
- Explore ways to understand human ability to identify faces

Approach

- Decompose faces to eigenface features



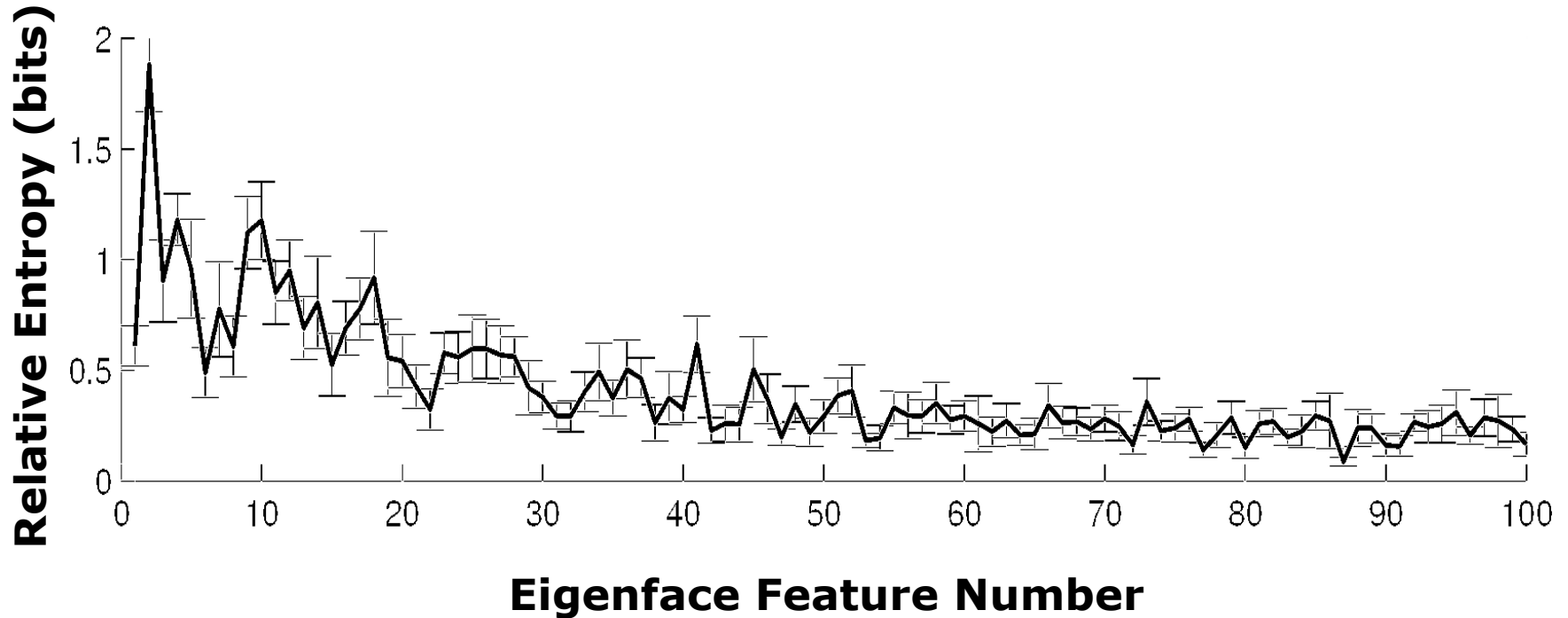
- Fit features to Gaussian
 - Mean $\boldsymbol{\mu}_q, \boldsymbol{\mu}_p$
 - Covariance $\boldsymbol{\Sigma}_q, \boldsymbol{\Sigma}_p$
- Calculate $D(p||q)$ for each individual
- Average $D(p||q)$ for all individuals

Face Database

- Used Aberdeen Face database
 - 18 frontal images of 16 persons
 - Variability in lighting and expression between images

- Correlation calculations are limited
 - Given 18 images, Σ_p is non-singular for more than 17 features
 - Don't calculate overall information

Average Information per feature



- ❑ Most informative feature: #2
- ❑ Decreasing information with EF number

Applications: *biometric*

- Meta algorithm
 - Evaluate a new biometric feature
- Biometric Performance limits
 - Template size limits
 - Inherent match performance limits
- Feasibility of Biometric Encryption
 - Limits to Key Length

Applications: *abstract*

- Quantify privacy
 - What is the privacy risk due to the release of certain information?
 - What is the privacy gain in obscuring faces?
- Uniqueness of biometrics
 - Approach to address: "Are faces / fingerprints / irises unique?"

Issues with biometric features

- Calculation requires feature representation
- Feature dimension may not be constant
 - Fingerprint minutiae
- Raw images must be aligned and scaled
 - Errors in pre-processing will affect measure

Issues with Entropy measure

- Features are correlated
 - e.g. height and weight
 - Multiple measures of same underlying feature (PCA vs. ICA Face Rec.)
- Feature distributions are complex
 - Gaussian model used is not correct for minutiae angle
- Lots of data required to model
 - Currently researching approaches using less data

Conclusions

- Approach to measuring information content of a biometric system
- Relative Entropy is appropriate measure
- Help explain *legal, social, performance* issues