Engineering Privacy in Public

James Alexander and Jonathan Smith
University of Pennsylvania

Introduction

- Project Goal: A generalized, experimentally-validated privacy metric
- First experiment: Defeating face recognition
- Experiments with more biometrics to follow

Talk Overview

- Project Goals
- Face Recognition: Methodology and Evaluation
- Disguise Slide Show
- Analysis
- Future Work

Value of PET Generality

- Though details differ wildly, the goal of all PETs is the same: to help the user *not* be identified
- Advantages of a common framework:
  - User can tell where they get the most “bang for the buck”
  - Easier to evaluate the combination of severals PETs in the presence of multimode surveillance

Project Goal

To develop a “benefit” metric for evaluation of privacy enhancing technologies

- Propose candidate metrics and evaluate against empirically-measured PET performance

General Properties

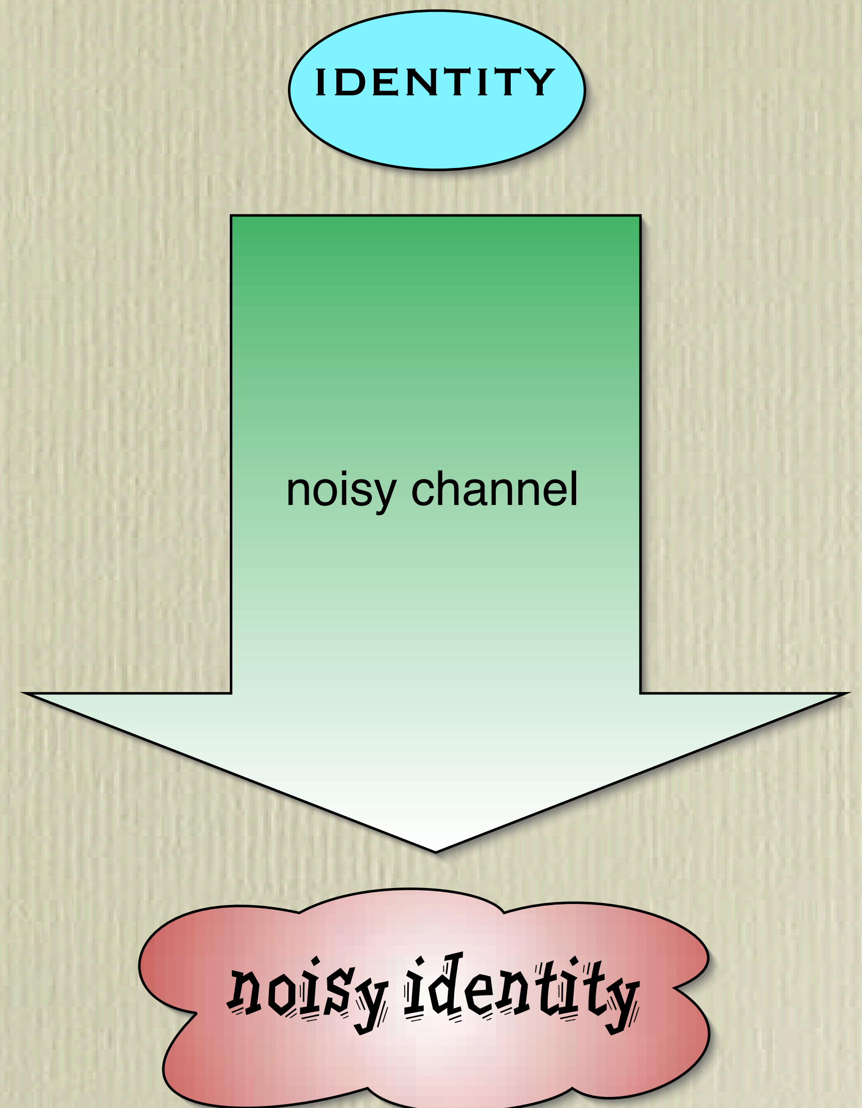
- Suitable for cost / benefit analysis regardless of how cost is quantified
- Explainable to a lay person
- Places reliable bounds on how well an adversary can do, even without precise knowledge of adversary's methods

Modeling Privacy

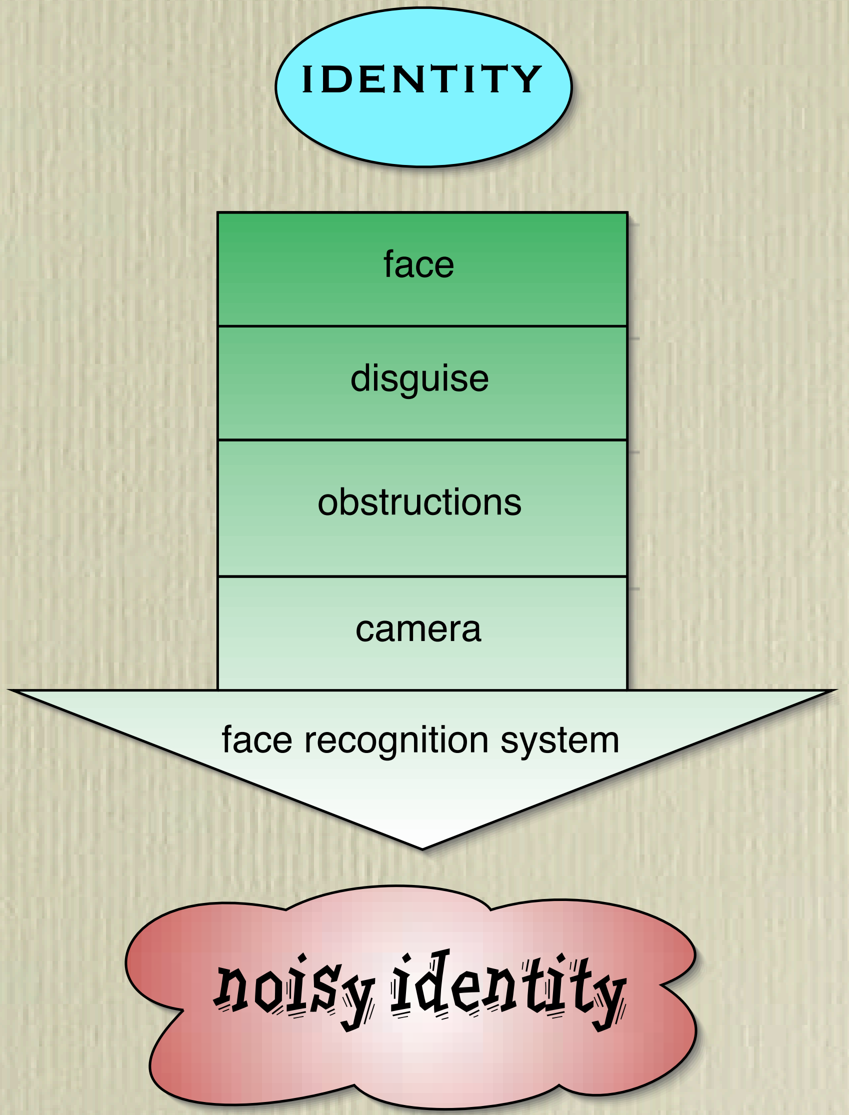
Adversary knows some predicate holds of a particular individual

- He builds a probability distribution of this predicate over the set of all individuals
- Job of a PET is to make sure the correct individual does not stand out in the distribution

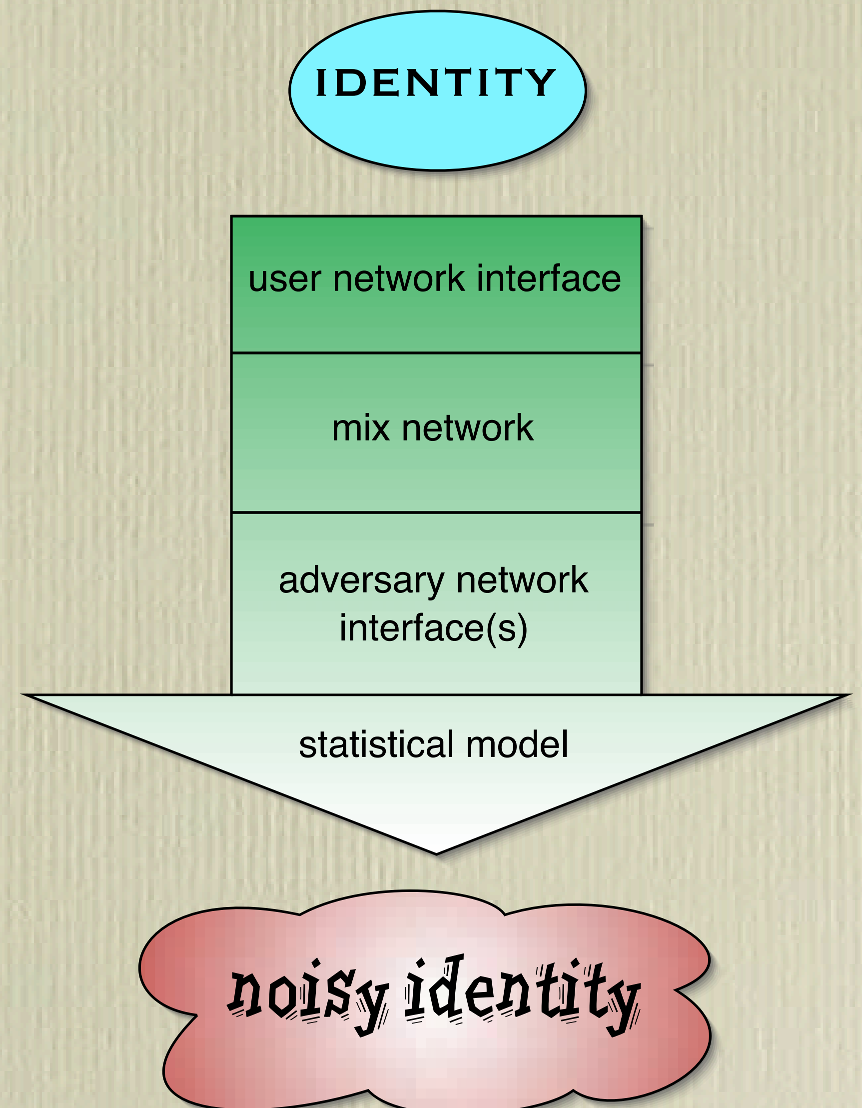
Privacy as a Noisy Channel



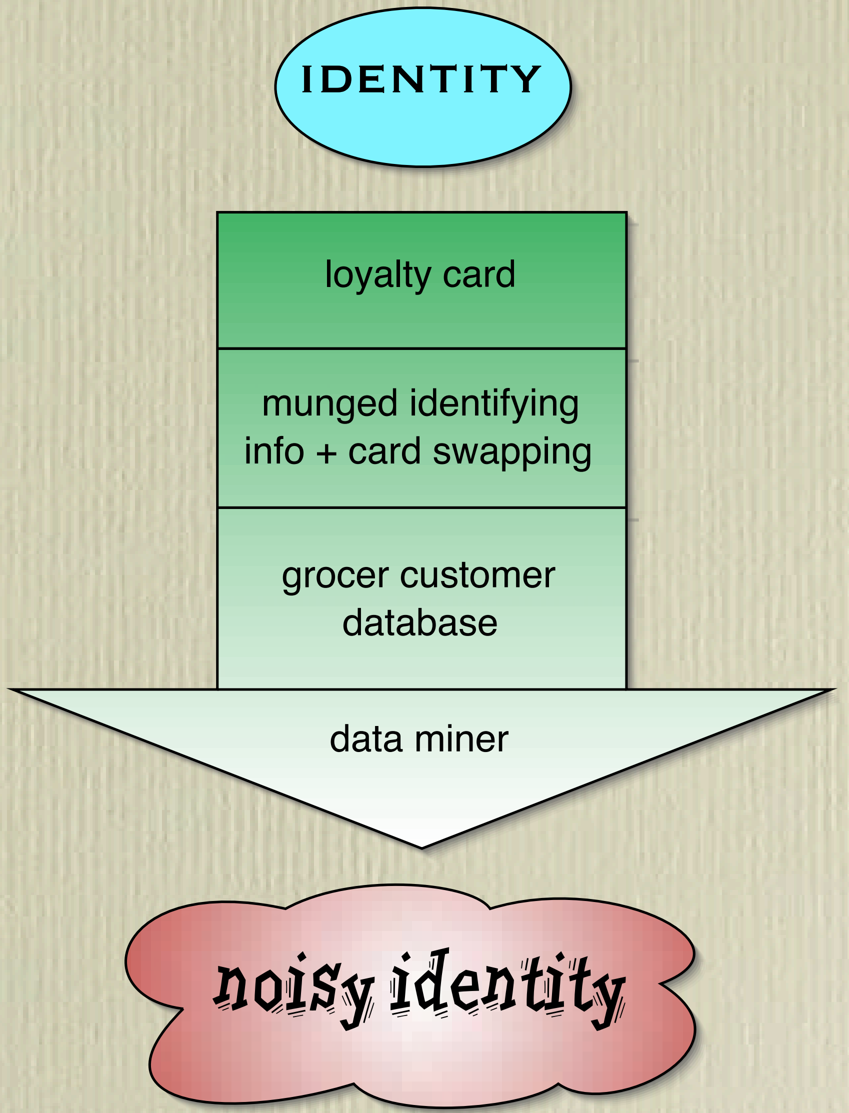
Face Recognition



Mix Networks



Store Loyalty Cards



Challenges

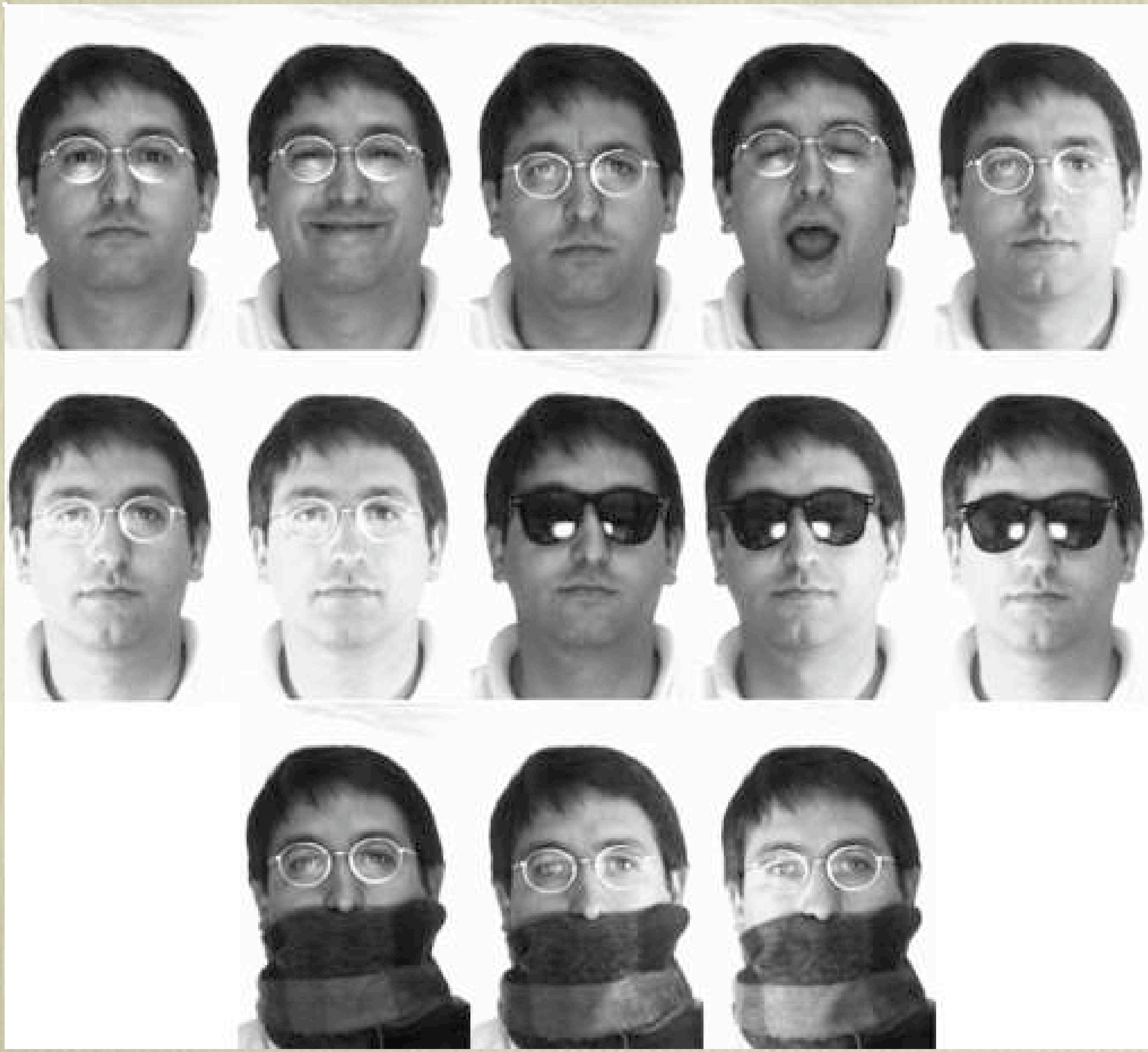
- We want to predict entropy in the adversary's model - we can't measure it directly, but perhaps can place bounds on it
- Theory of non-cooperating communicators is not well-explored
    - What are the limits of a communication channel employing a sabotaged encoding?
    - What if noise sources are not random?

Methodolgy

- Tested face recognition system an eigenfaces system used in the FERET evaluation
- 3816 FERET images used as distractors
- New pictures added to match FERET specs
- Facial occlusion images from AR database give statistical behavior of two particular disguises

Sample Baselines

AR Sample



Adversary Model

- Can obtain high-quality frontal probe images
- Might have more than one gallery image of you
- System output will consist of up to N candidate matches, presented to an operator for confirmation
- Face recognition system will be deployed on a large scale
- Do not know if a minimum likelihood cut-off used

Score Function

$$w_x(i) = \begin{cases} N - i + 1 & \text{if the candidate in the } i\text{th position} \\ & \text{is really } x \text{ (i.e. a match)} \\ 0 & \text{otherwise} \end{cases}$$

$$\text{score}(x) = \frac{\sum_{i=1}^{N} w_x(i)}{\sum_{i=1}^{N} i}$$

Effective Disguises



AR performance

Image group	Accuracy	Mean Score
baseline	99.6%	0.6947
sunglasses	15.0%	0.0344
scarf	58.7%	0.2323
overall	45.8%	0.2136





A minor difficulty

- Problem: The score function doesn't allow performance comparison among disguises that all score zero
- Solution: Morphs!



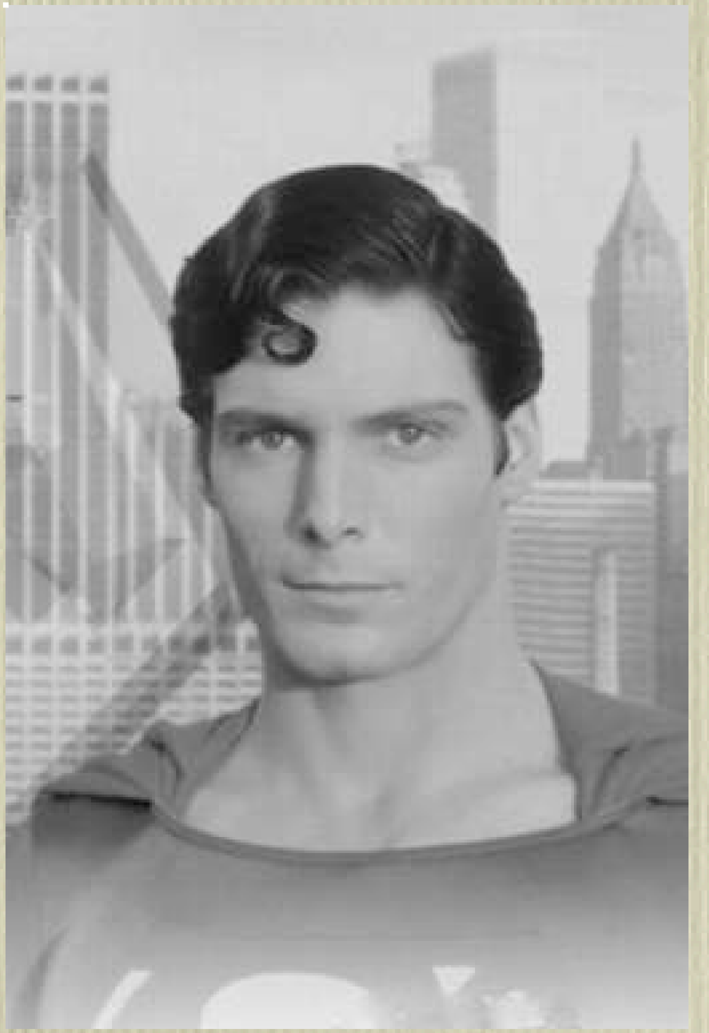








Ineffective Disguises







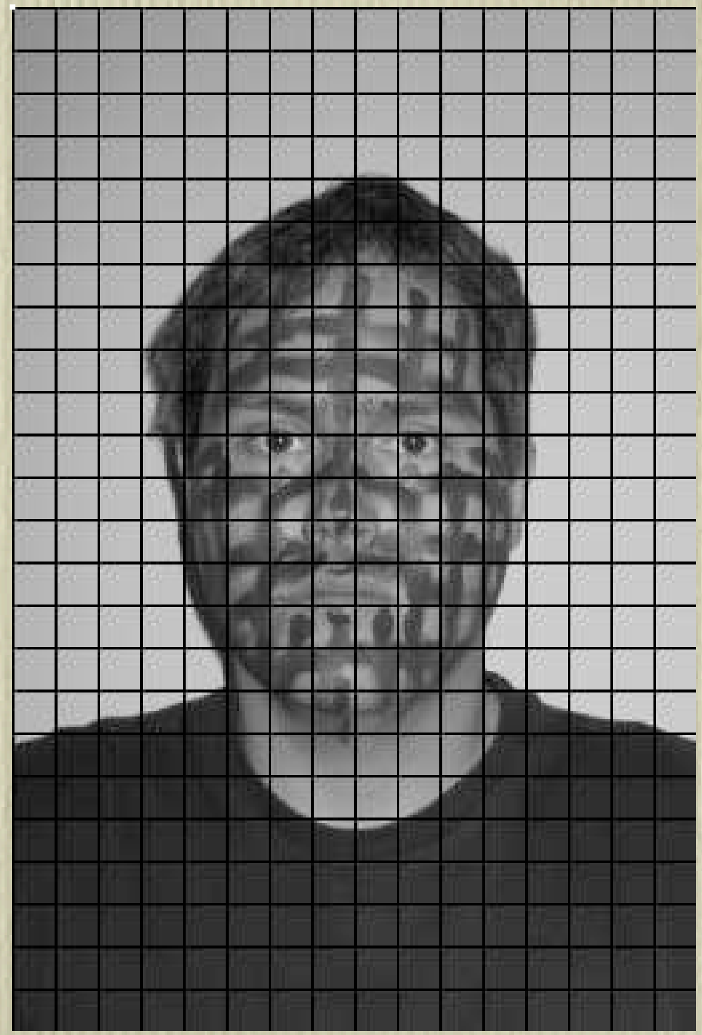
What's going on?

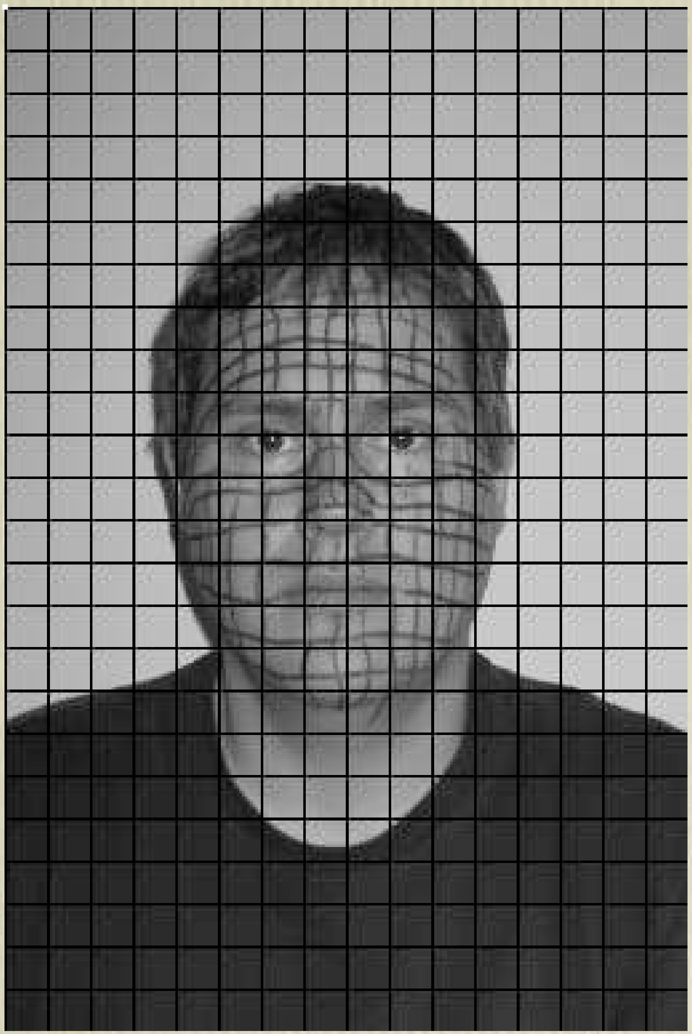
- The system is attempting to match facial features and their positions to the closest matches in its training data
- To fool it, we need to obscure or remove existing features, or provide decoy features for it to find
- Features are composed of *contrasts* in the photographic data



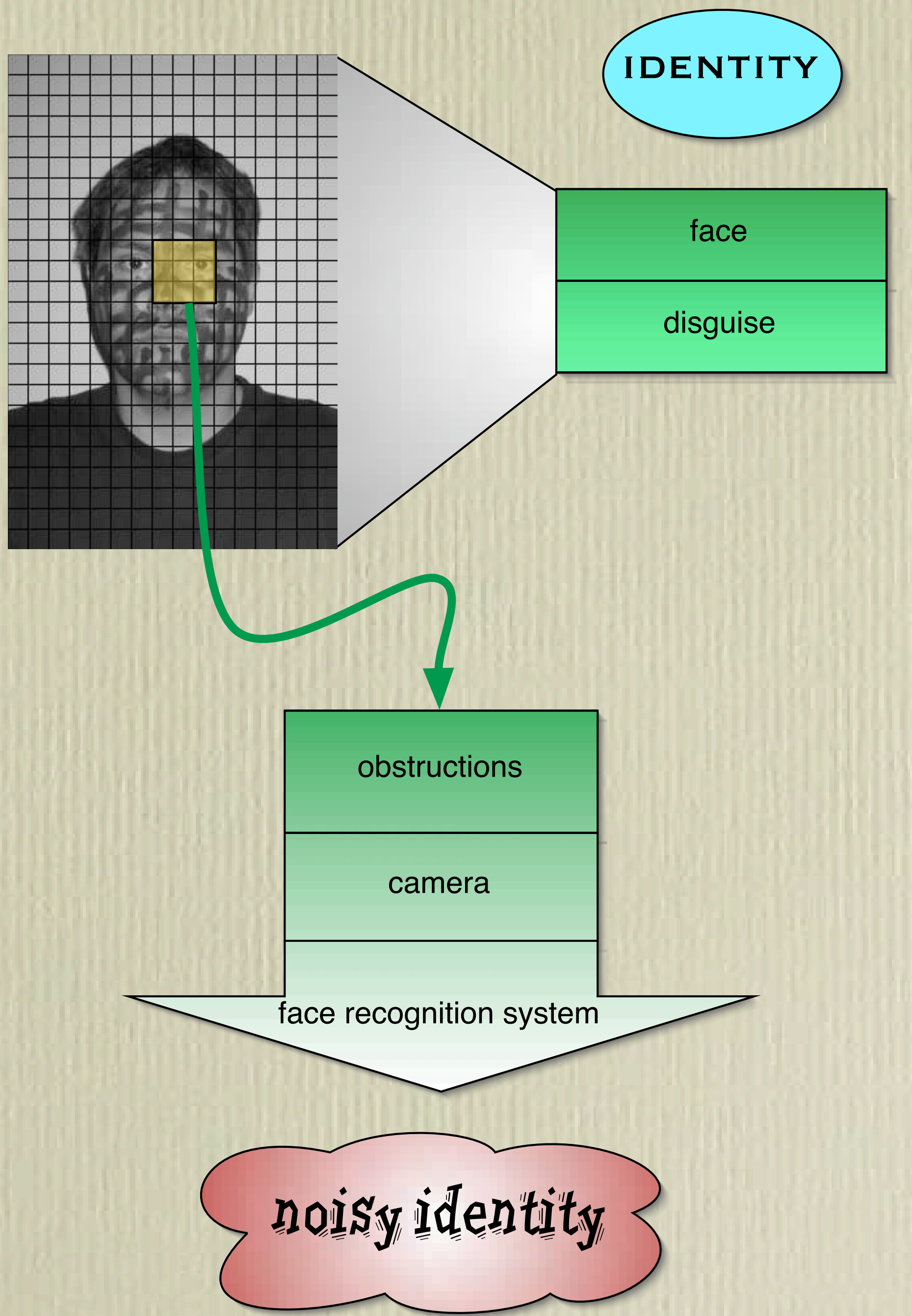



Grid Model





A Grid in the Noisy Channel



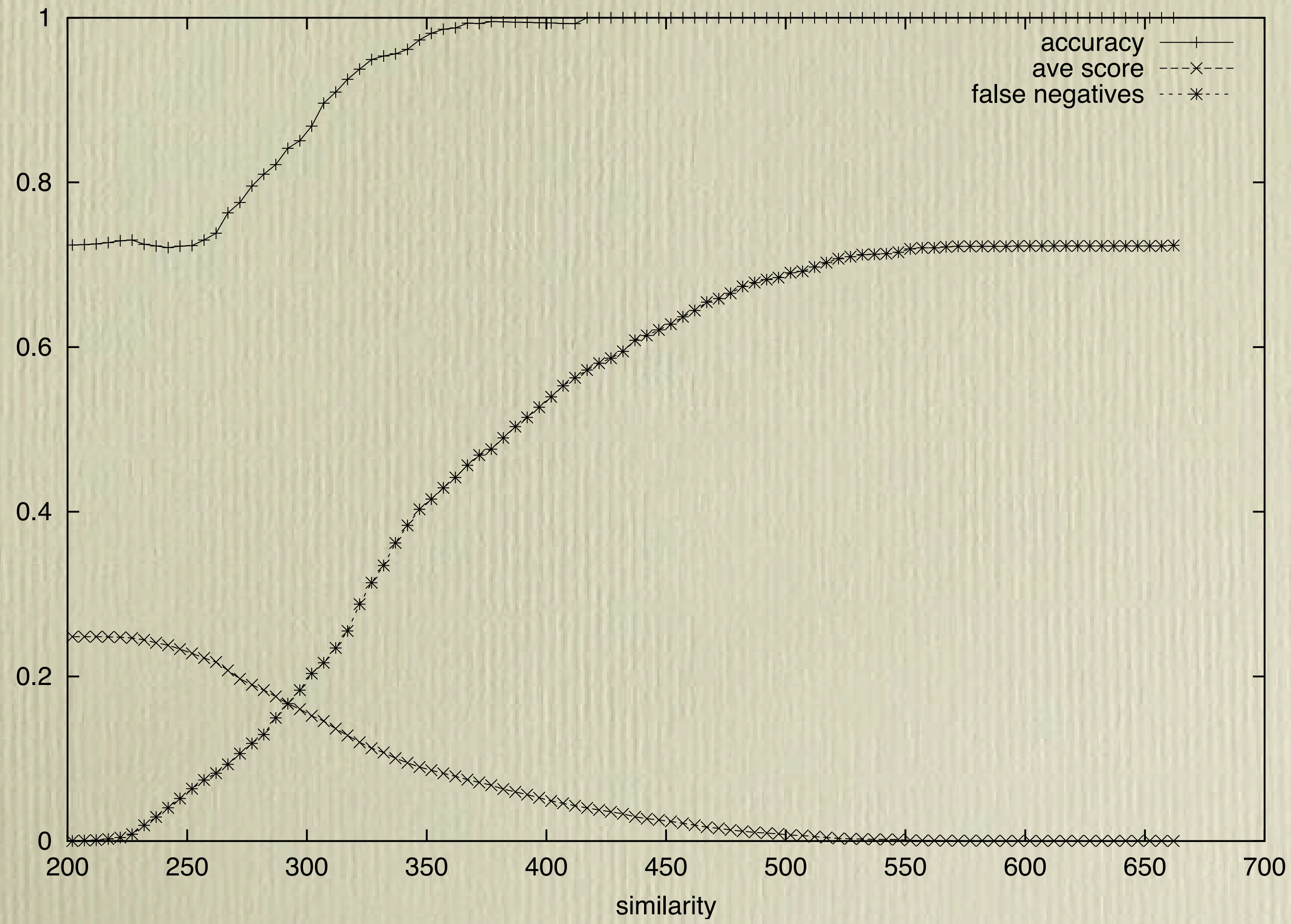
Refining the Grid

Experiments in progress in order to determine:

- The critical size separating features from non-features (i.e. the right size of grid boxes)
- The weights representing the differing importance of each grid position to system performance

An anomaly

Performance Trade-offs



Future Work

- Elaborate the grid model further
- Test disguises on more subjects
- Replicate with a face recognition system with a very different underlying model (e.g. FaceIt)
- Extend framework to more biometrics, and beyond

Questions?