UCCS
University of Colorado at Colorado Springs

A Fusion-Based Approach To Enhancing Multi-Modal Biometric Recognition System Failure Prediction and Overall Performance

Walter J. Scheirer and Terrance E. Boult
VAST Lab University of Colorado at Colorado Springs and Securics, Inc

Securics®
The science of security with the protection of privacy™

What is *Failure* in a biometric recognition system?

Statistical Type I Error

- An imposter has achieved a match within the gallery

Statistical Type II Error

- A probe fails to match out of the top n scores for rank n recognition

Why is failure prediction important?

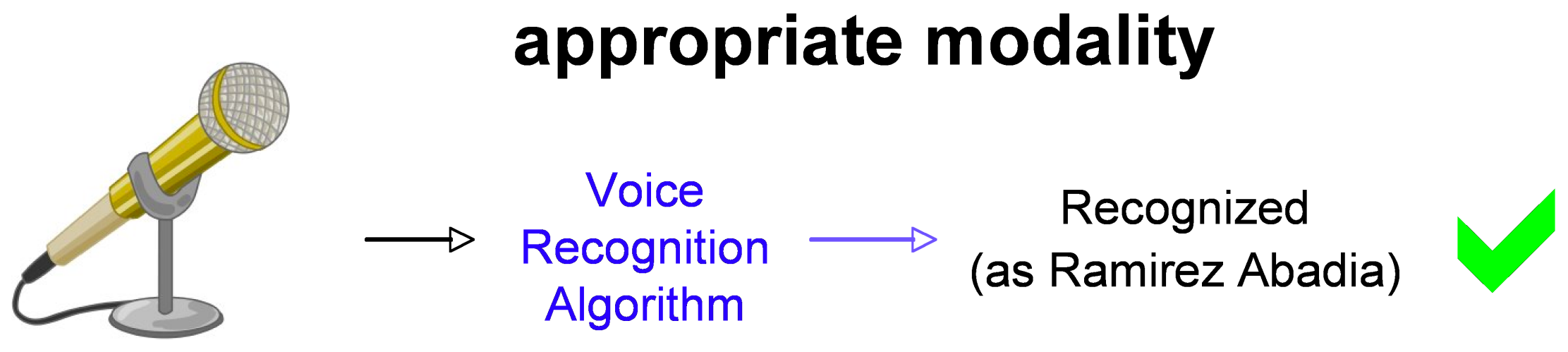
• Per instance failure prediction is critical for sensitive installations, screening areas, and surveillance posts

• The recent case of Columbian drug cartel leader Juan Carlos Ramirez Abadia highlights the need for failure prediction in biometric recognition. Ramirez Abadia underwent plastic surgery to evade facial recognition, but was apprehended with the aid of voice recognition.

Face Recognition Algorithm
Rejected (not in DB)

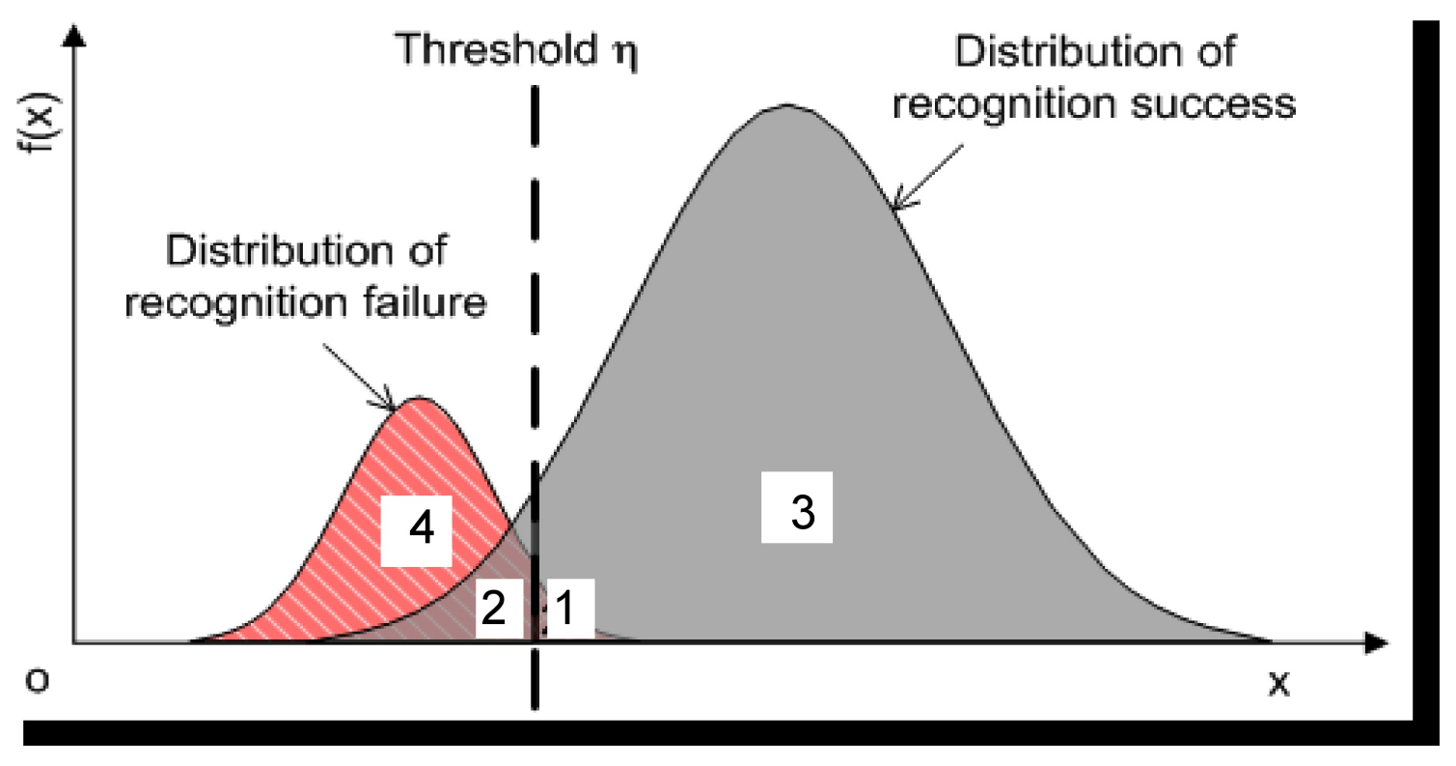
Ramirez Abadia

Prediction can choose the appropriate modality



Failure Prediction

Threshold a per datum reliability measure to predict recognition system success produces 4 different "cases"



Case 1: False Accept, prediction of success contrary to ground-truth

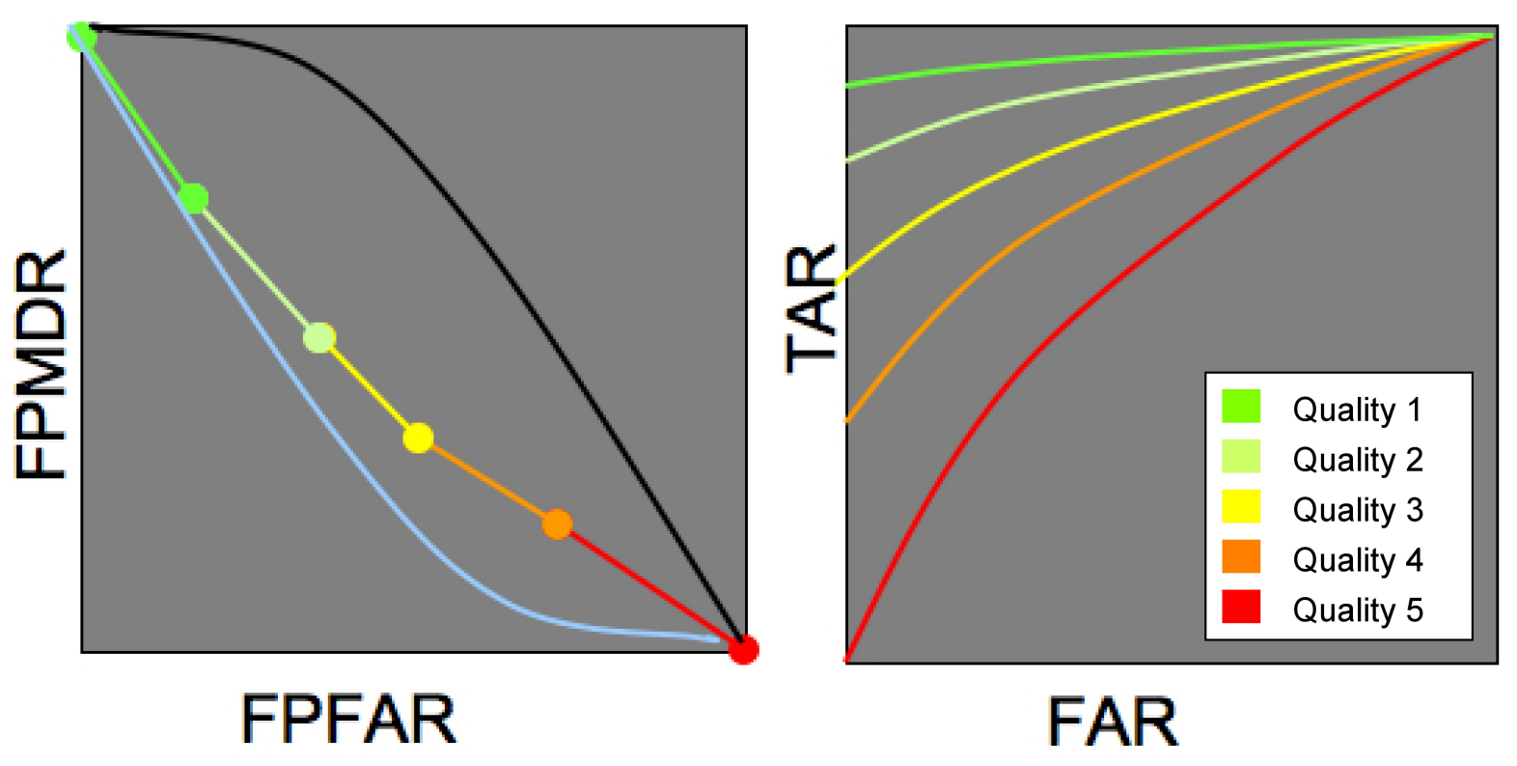
Case 2: False Accept, prediction of failure to the contrary of ground-truth

Case 3: correct prediction of success

Case 4: correct prediction of failure

The FPROC curve is defined by the FPFAR and FPMDR

$$\mathrm{FPFAR} = \frac{|\mathrm{Case}\ 2|}{|\mathrm{Case}\ 2| + |\mathrm{Case}\ 3|} \qquad \mathrm{FPMDR} = \frac{|\mathrm{Case}\ 1|}{|\mathrm{Case}\ 1| + |\mathrm{Case}\ 4|}$$



FPROC vs. CMC. Segmenting the data on quality inflates the difference. Using full data sets in the FPROC allows us to vary the quality threshold.

Previous Work

• W. Scheirer, A. Bendale, and T. Boult, "Predicting Biometric Facial Recognition Failure with Similarity Surfaces and Support Vector Machines," in *Proc. of the IEEE Computer Society Workshop on Biometrics,* 2008

• B. Xie, V. Ramesh, Y. Zhu, and T. Boult, "On Channel Reliability Measure Training for Multi-Camera Face Recognition," in *Proc. of the IEEE Workshop on the Applications of Computer Vision*, 2007

• T. Riopka and T. Boult, "Classification Enhancement via Biometric Pattern Perturbation," in *Proc. of the IAPR Conference on Audio- and Video-Based Biometric Person Authentication*, 2005

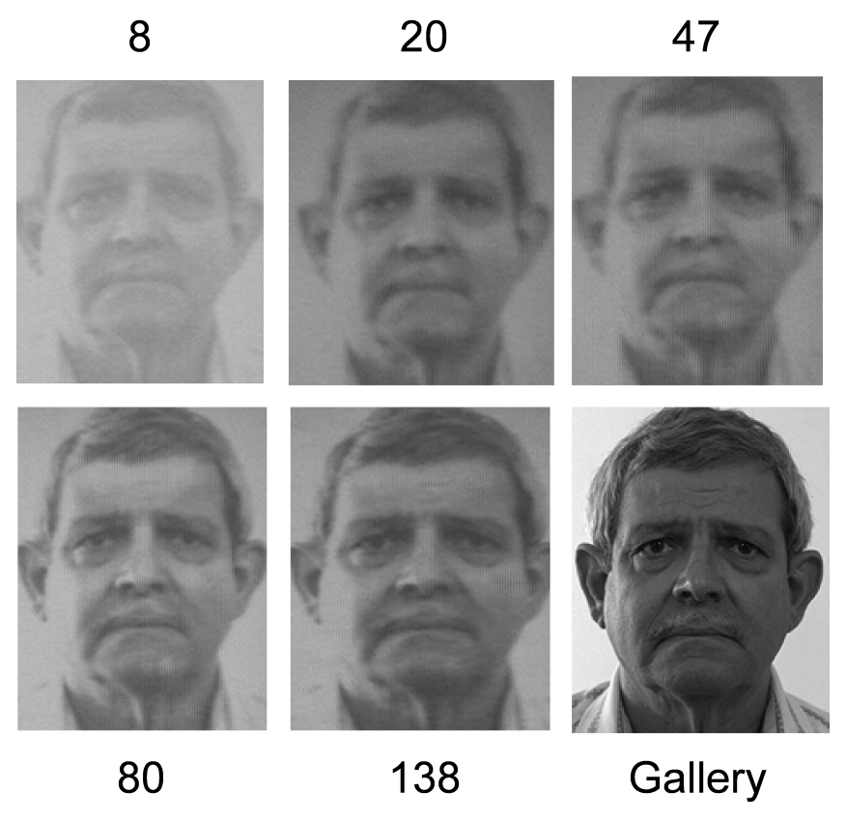
• W. Li, X. Gao, and T. Boult, "Predicting Biometric System Failure," in Proc. of the *IEEE Conference on Computational Intelligence for Homeland Security and Personal Safety*, 2005

This work was supported in part by DHS SBIR NBCH080054 and NSF PFI Award Number 0650251

Why not just use image quality as a predictor?

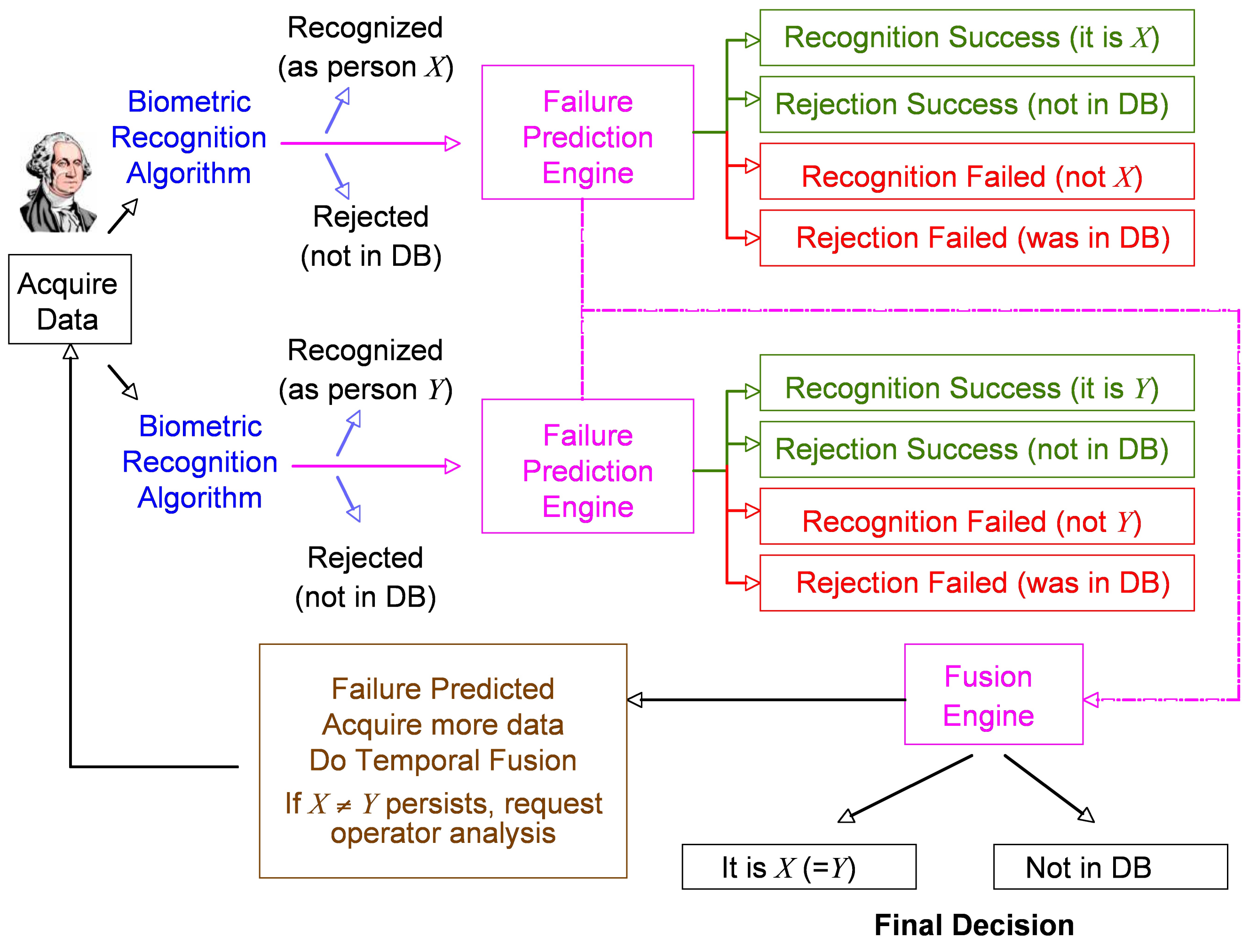
"Quality is not in the eye of the beholder; it is in the recognition performance figures!"[1]

• Quality, while a solid predictor overall, can sometimes be misleading for "per instance" failure prediction.



5 images of varying quality, and associated rank scores, along with the original gallery image for comparison. Apparent quality is not always correlated with rank!

The process flow of a of a multi-modal recognition system incorporating failure prediction based fusion


Acquire Data
Biometric Recognition Algorithm
Recognized (as person X)
Rejected (not in DB)
Failure Prediction Engine
Recognition Success (it is X)
Rejection Success (not in DB)
Recognition Failed (not X)
Rejection Failed (was in DB)
Biometric Recognition Algorithm
Recognized (as person Y)
Rejected (not in DB)
Failure Prediction Engine
Recognition Success (it is Y)
Rejection Success (not in DB)
Recognition Failed (not Y)
Rejection Failed (was in DB)
Fusion Engine
Failure Predicted
Acquire more data
Do Temporal Fusion
If $X \neq Y$ persists, request operator analysis
It is X (=Y)
Not in DB
Final Decision

Features for Failure Prediction

For all features: sort all distance measurements or similarity scores from best to worst, take the minimum of minimums over all views for each gallery entry, then consider the top k scores for feature vector generation.

• $\boldsymbol{\Delta_{i,j\ldots k}}$ defined as ((sorted score i) - (sorted score j), (sorted score i) - (sorted score $j+1$), . . ., (sorted i) - (sorted score k)), where $j = i + 1$. Feature vectors may vary in length, as a function of the index i.

• **DCT** coefficients produced from the top n scores

Fusion Techniques

Fusion across features, algorithms, and modalities increases our chance of correctly predicting failure.

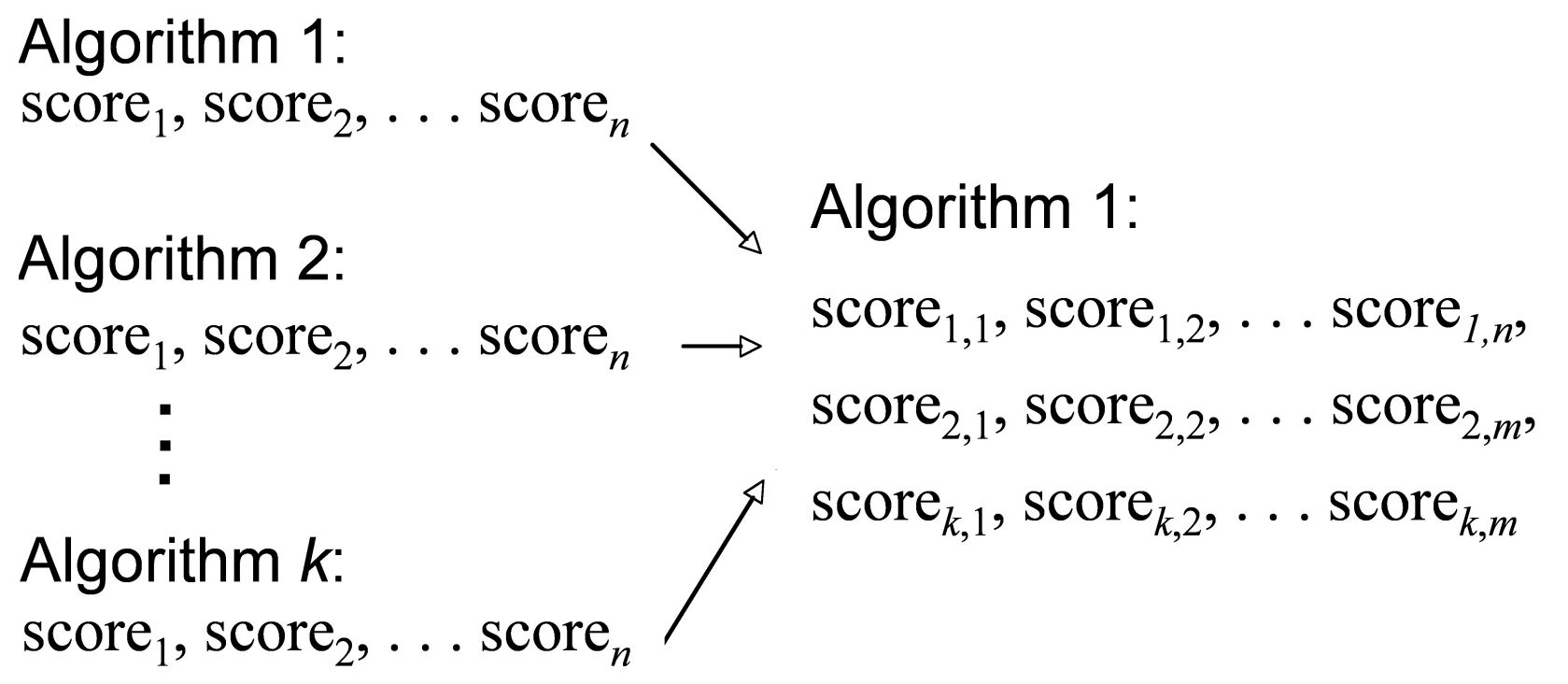
Threshold over all decisions across all features:

$$T\left(\begin{matrix} d(\mathrm{feature}_1) \\ d(\mathrm{feature}_2) \\ \ldots \\ d(\mathrm{feature}_n) \end{matrix}\right)$$

individual thresholds across all decisions across features:

$$\begin{matrix} T(d(\mathrm{feature}_1)) \\ T(d(\mathrm{feature}_2)) \\ \vdots \\ T(d(\mathrm{feature}_n)) \end{matrix}$$

Combine data from one or more algorithms in another algorithms:


Algorithm 1:
$\mathrm{score}_1, \mathrm{score}_2, \ldots \mathrm{score}_n$
Algorithm 2:
$\mathrm{score}_1, \mathrm{score}_2, \ldots \mathrm{score}_n$
⋮
Algorithm k:
$\mathrm{score}_1, \mathrm{score}_2, \ldots \mathrm{score}_n$
Algorithm 1:
$\mathrm{score}_{1,1}, \mathrm{score}_{1,2}, \ldots \mathrm{score}_{1,n},$
$\mathrm{score}_{2,1}, \mathrm{score}_{2,2}, \ldots \mathrm{score}_{2,m},$
$\mathrm{score}_{k,1}, \mathrm{score}_{k,2}, \ldots \mathrm{score}_{k,m}$

[1] P. Flynn, "Ice Mining: Quality and Demographic Investigations of Ice 2006 Performance Results," Presentation at the MBGC Kick-off Workshop, 2008

[2] National Institute of Standards and Technology, "NIST Biometric Score Set", 2004

[3] K. Nandakumar, Y. Chen, S. Dass, and A. Jain, "Likelihood Ratio Based Score Level Fusion,"in IEEE TPAMI, vol. 30, no. 2, pp. 342-347, 2008

[4] H. Korves, L. Nadel, B. Ulery, and D. Bevilacqua, "Multi-biometric Fusion: From Research to Operations," in *SIGMA*, Mitretek Systems, 2005

Computational Efficiency

The computational efficiency of this system (excluding the recognition system) is considered in two pieces: training and classification.

Training: To sort quickly gather the top k (never exceeding 10 in this work) scores out of n total scores, bucket sort can be used, requiring $O(n)$ operations. Computation for our best performing feature, $\Delta_{i,j\ldots k}$ is a simple series of linear operations (subtraction over a set of scores), and is thus $O(M)$ over M score series composed of the top k scores for each series. The offline training of a SVM is computationally expensive, with a time complexity of $O(M^3)$ over M training examples (feature vectors derived from the M score series). The complete time needed for training the system is $O(n + M + M^3)$ per classifier.

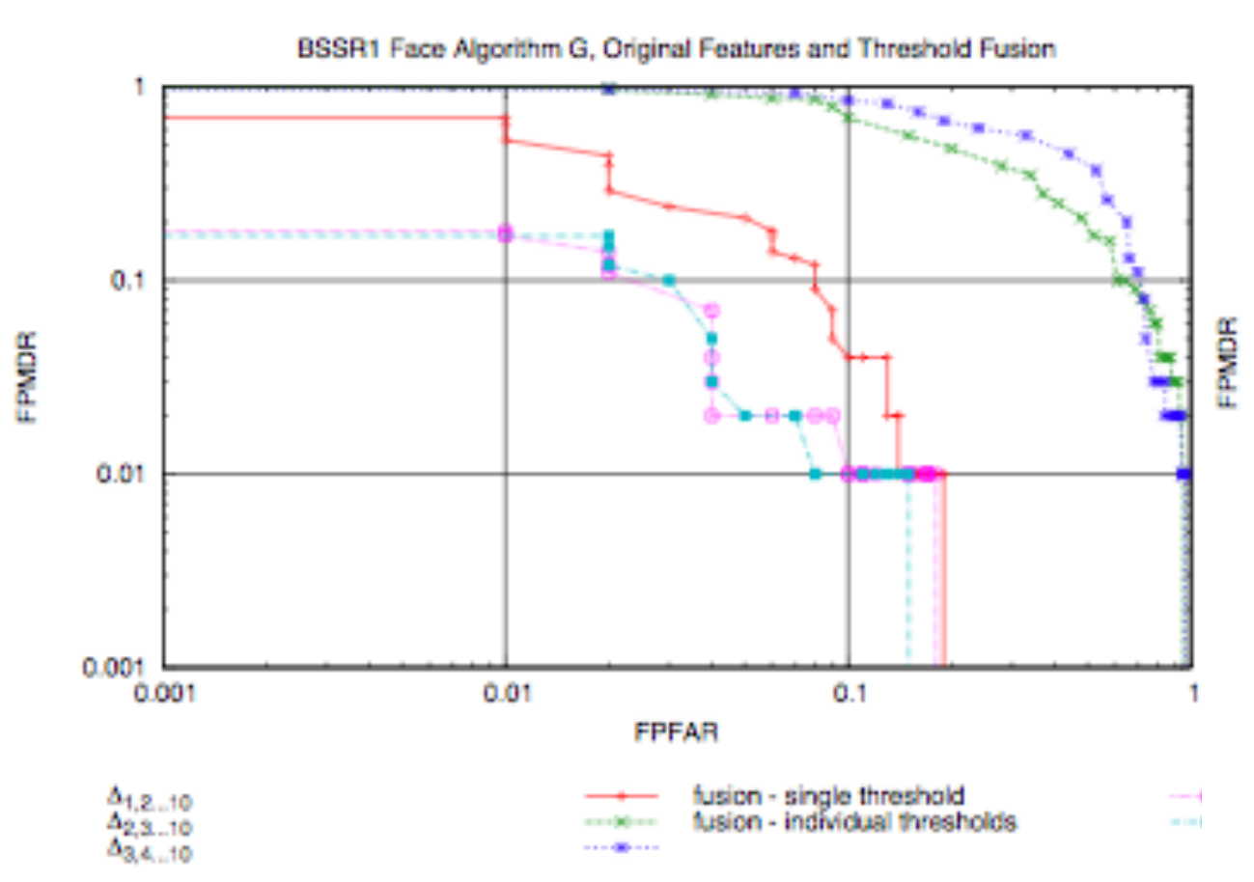
Classification: SVM classification is a linear operation of $O(M)$. The complete time needed for classification is $O(n + M)$ for both fusion before SVM classification and for fusion after SVM classification, where an extra pass over the SVM marginal distances is needed. This linear complexity is well suited for real time systems.

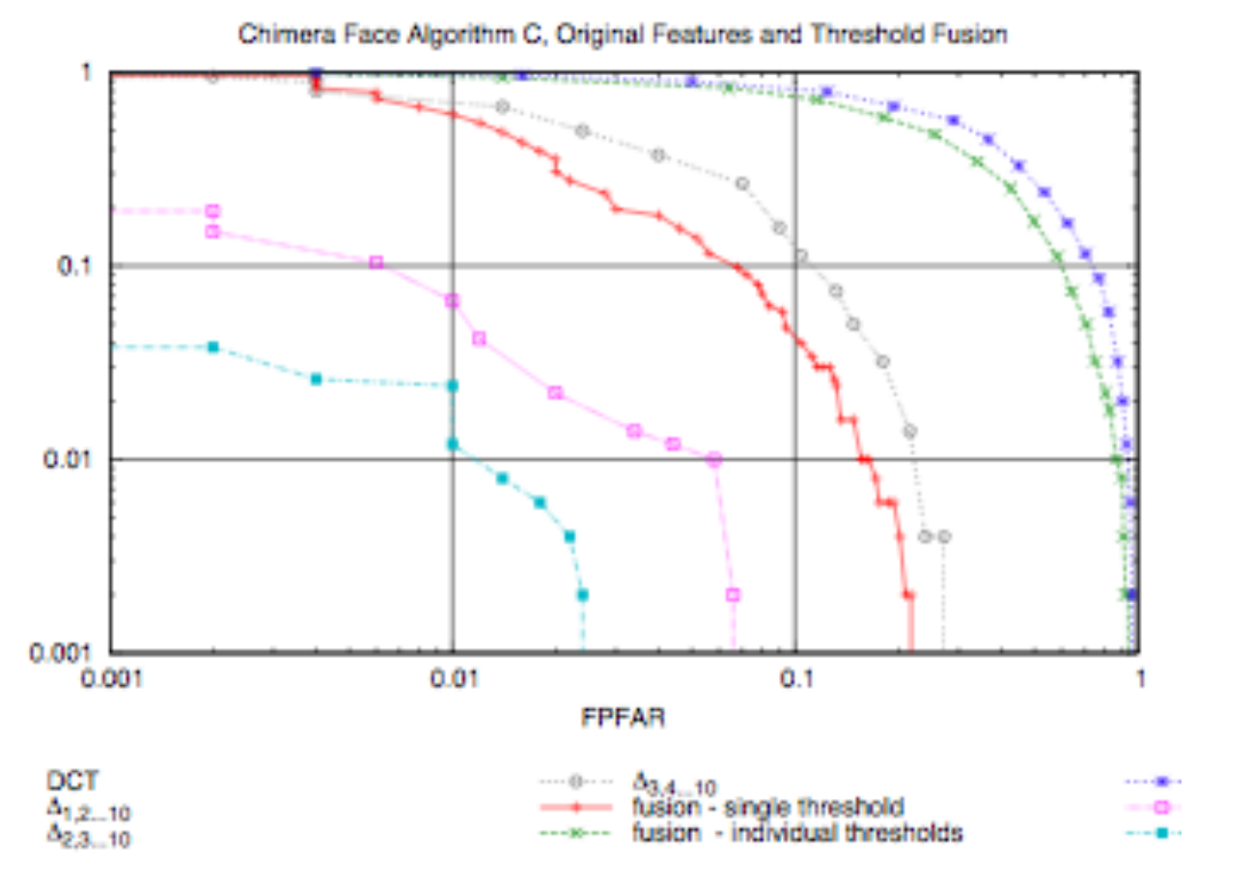
Experiments

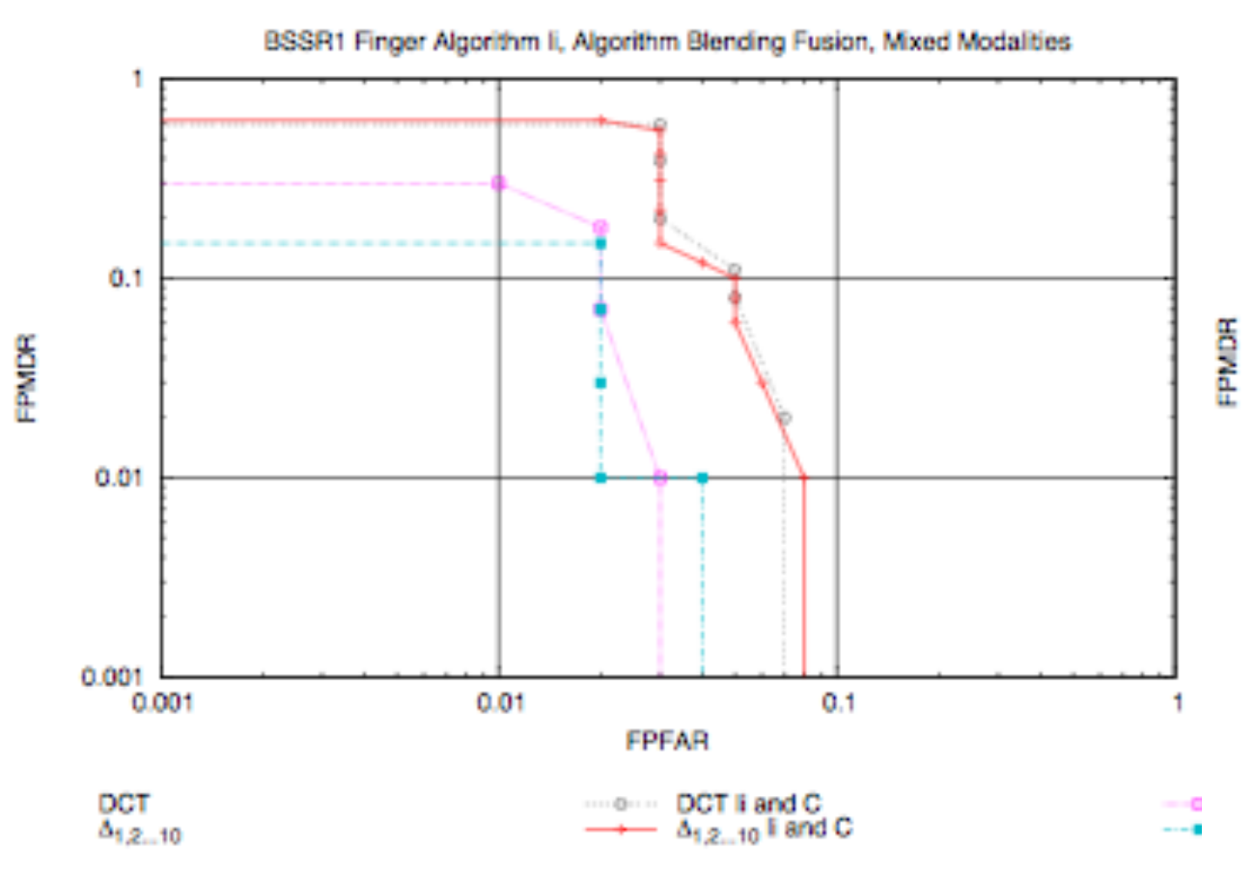
Data Set	Training Samples	Testing Samples	Face Algorithms	Finger Algorithms
BSSR1[2]	600	200	2	1
BSSR1 "Chimera"	6000	1000	2	1

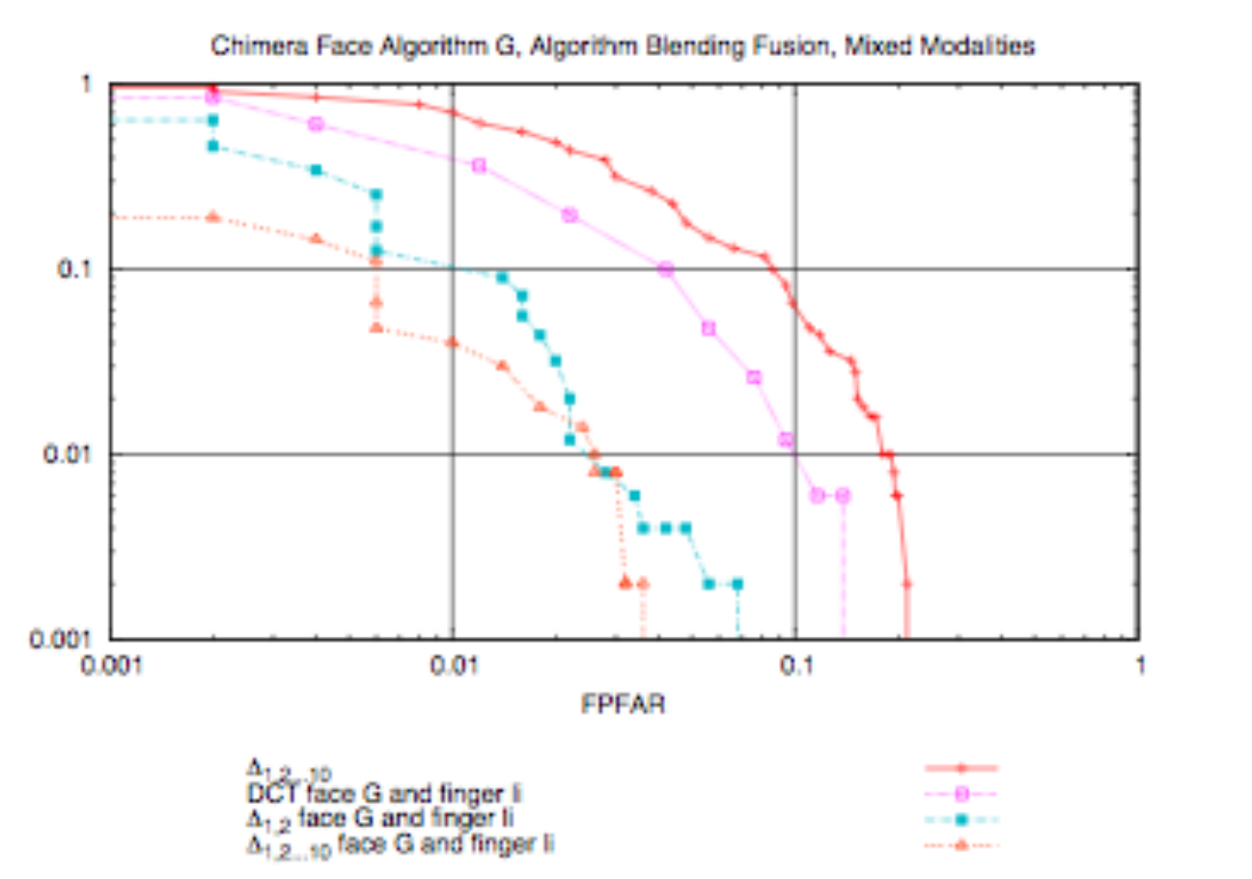
The data set breakdown for machine learning

The first set of experiments evaluates the performance of the fusion techniques over the baseline features for failure prediction. The expectation is that the fused prediction techniques will perform no worse than the original features, and in most cases, outperform them. *FPROC* curves follow.

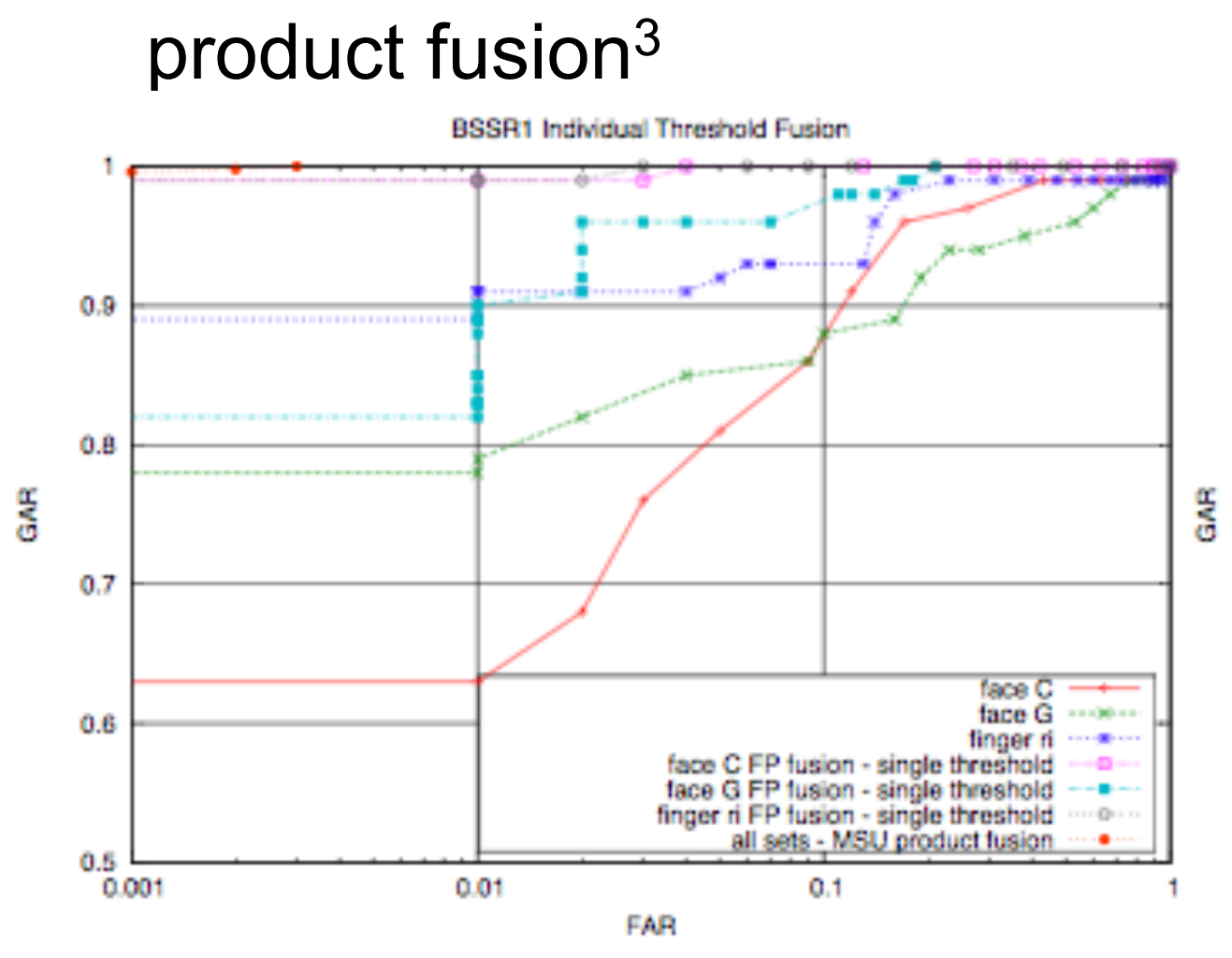


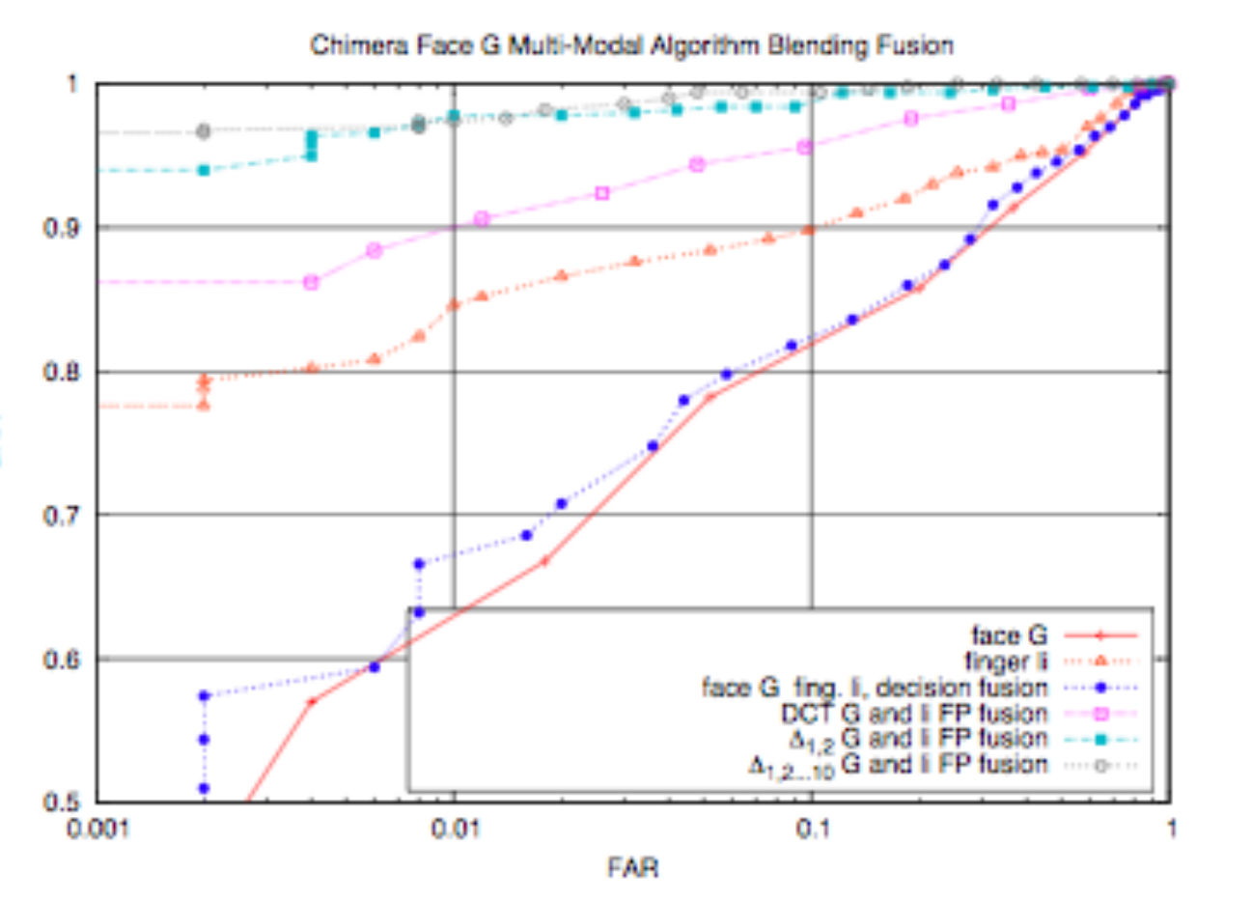


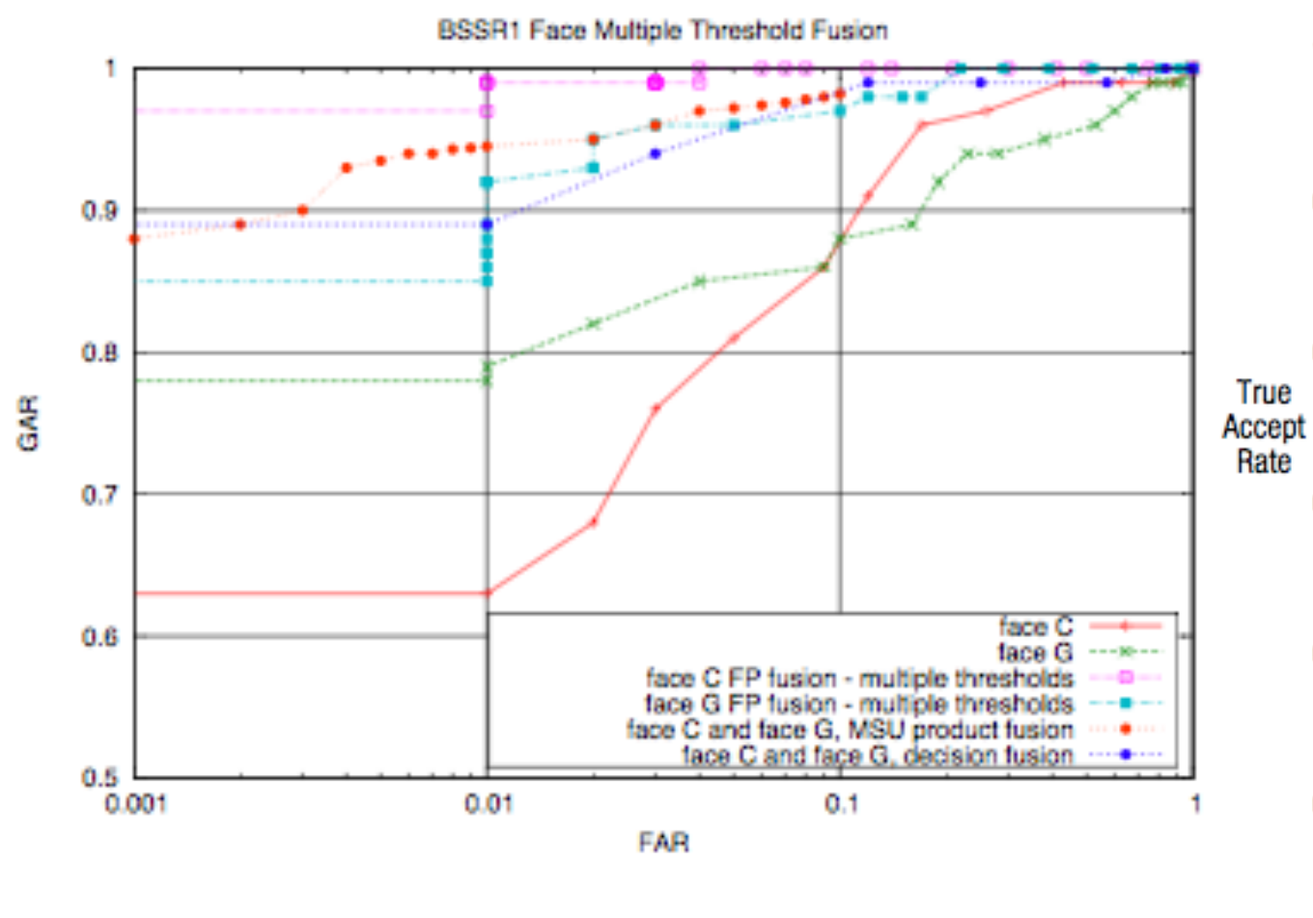


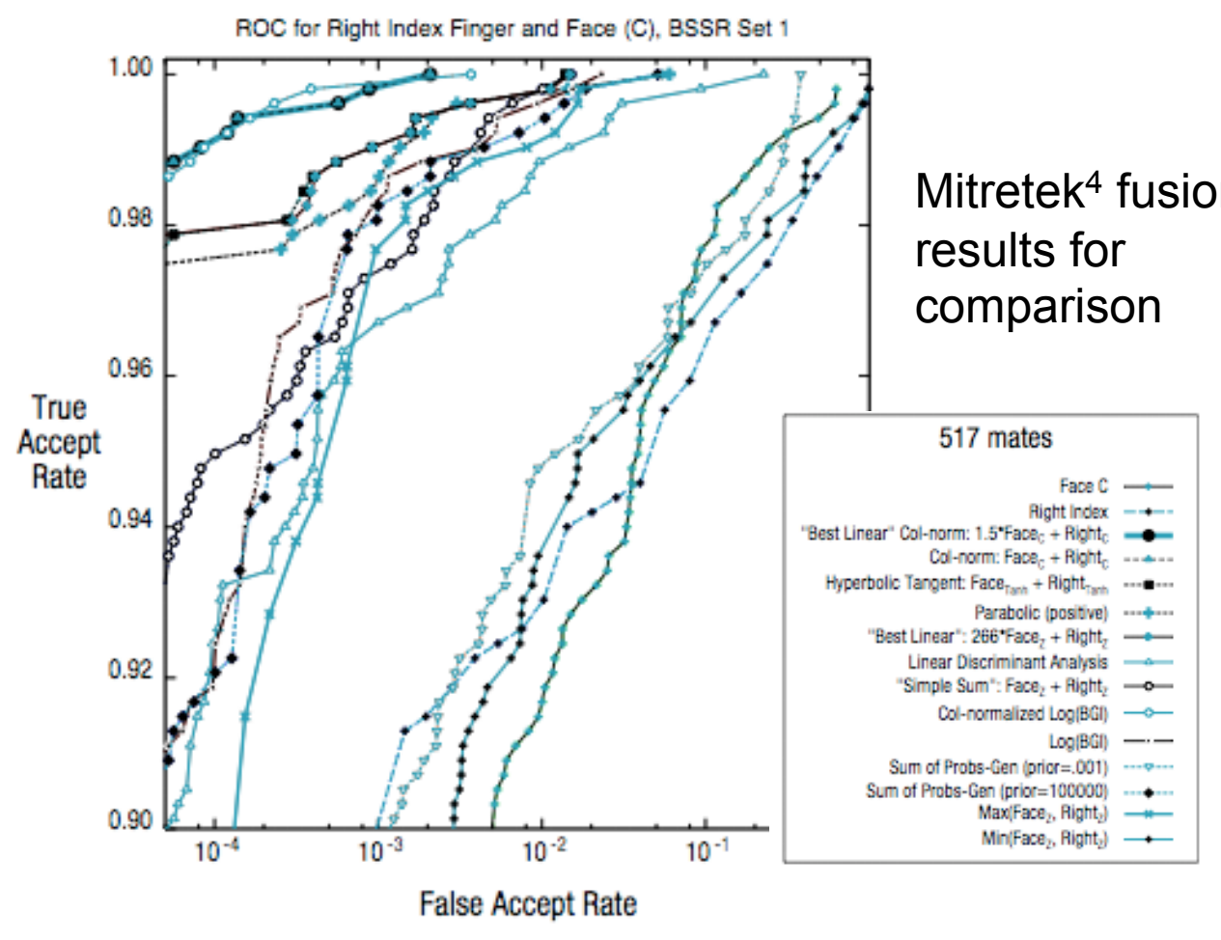


The second set of experiments was designed to evaluate the recognition system's performance after processing by the failure prediction fusion-based system. By predicting failure, we can apply further fusion to select results that have not failed. For the BSSR1 set, we compare our results to MSU's product fusion[3]









Mitretek[4] fusion results for comparison