Robust Low-Resolution Face Identification and Verification Using High-Resolution Features

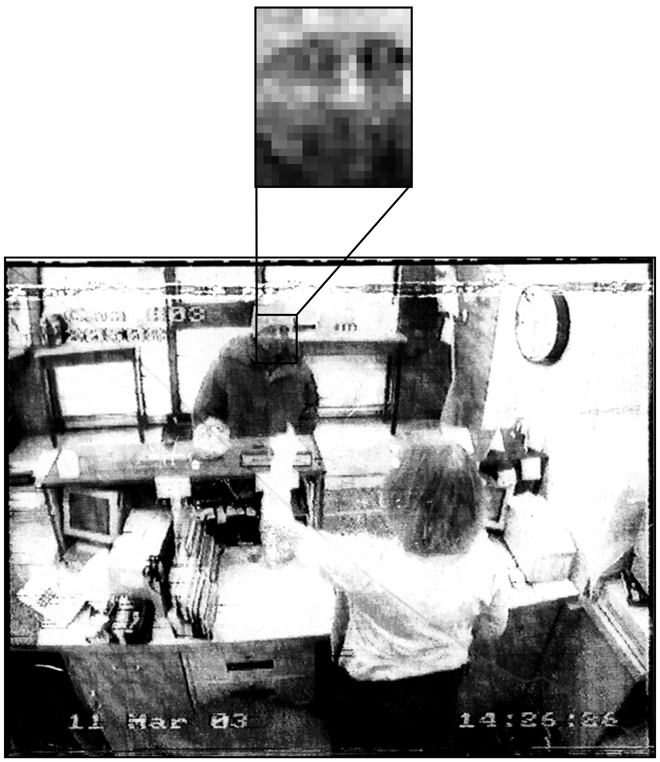
Dr. Pablo H. Hennings Yeomans[1], Prof. B.V.K. Vijaya Kumar[1], Dr. Simon Baker[2]

Electrical & Computer ENGINEERING

1. ECE Department, Carnegie Mellon University
2. Microsoft Research, Microsoft Corporation


CNN
OMNI
HOTEL

Surveillance and Forensics

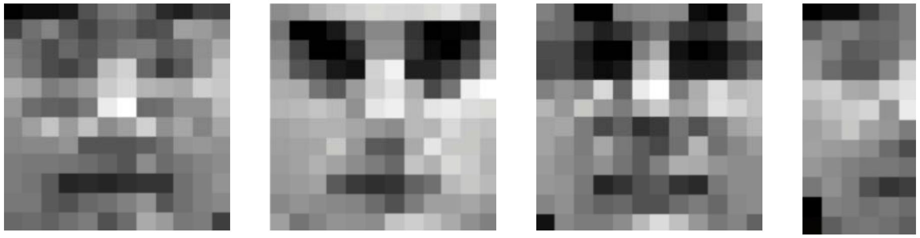


- In many surveillance scenarios people may be far from the camera and their faces may be small.
- Looking for suspects involves parsing through hundreds of hours of video.
- A terrorist crime was solved in Italy in 2002 by analysis of 52,000 hours of surveillance videos installed in rail stations.

Problem

- As media usage moves from images to video, low-resolution content is now more common than before.
- The problem from the biometrics perspective is the mismatch between training and testing resolutions.

Training Images



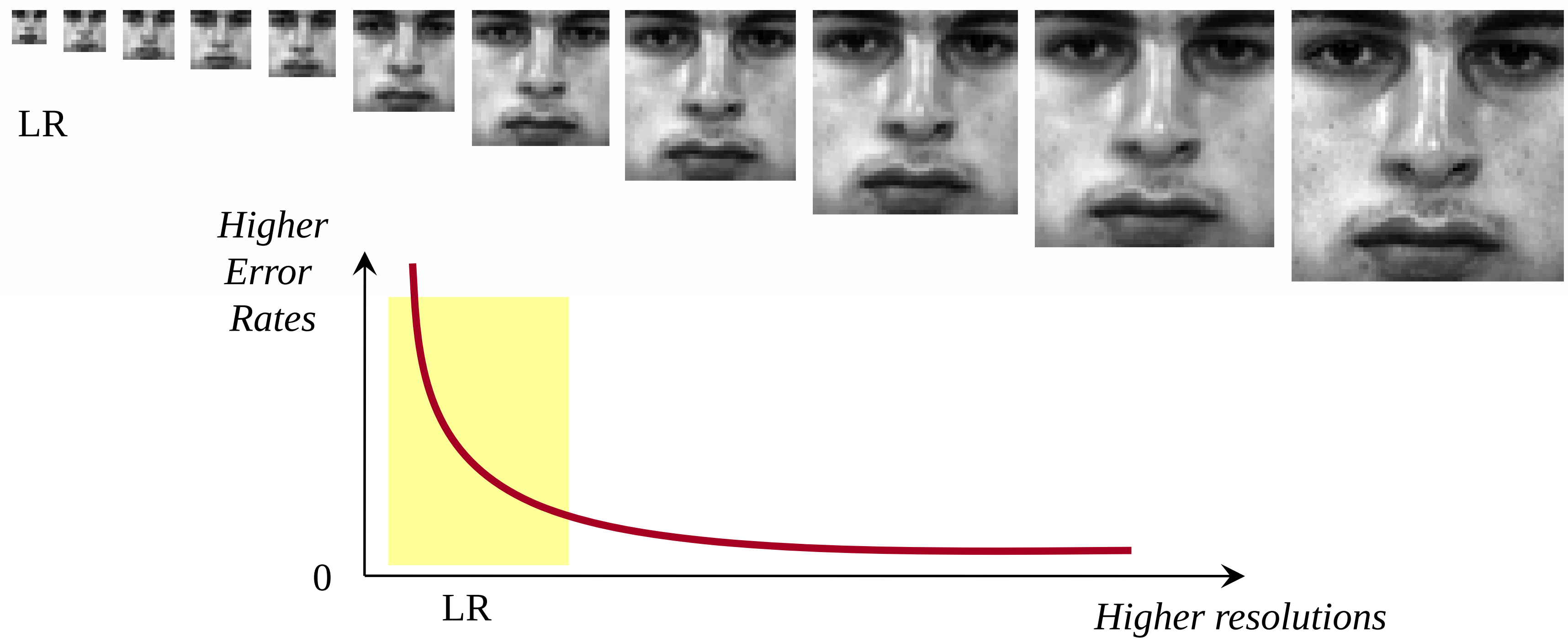
Probe Images

Outline

- Conventional approaches
- Proposed framework
- Generalizations that improve recognition
- Experiments and Results
- Conclusions

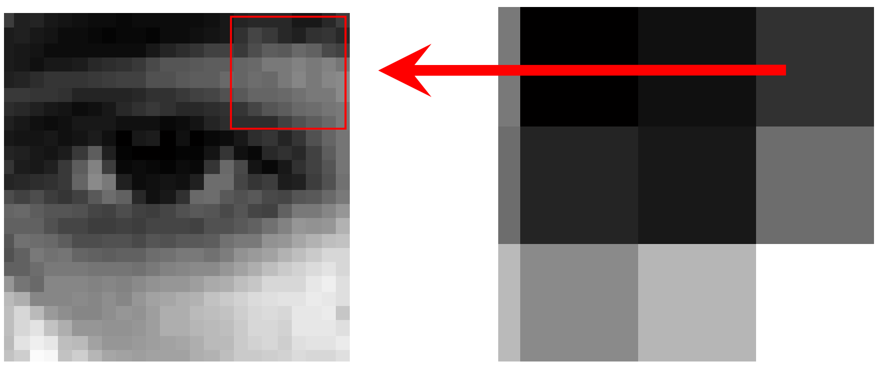
Face Recognition of Low-resolution Images

- Classification degrades considerably as the resolution of probe faces decreases to very low resolutions

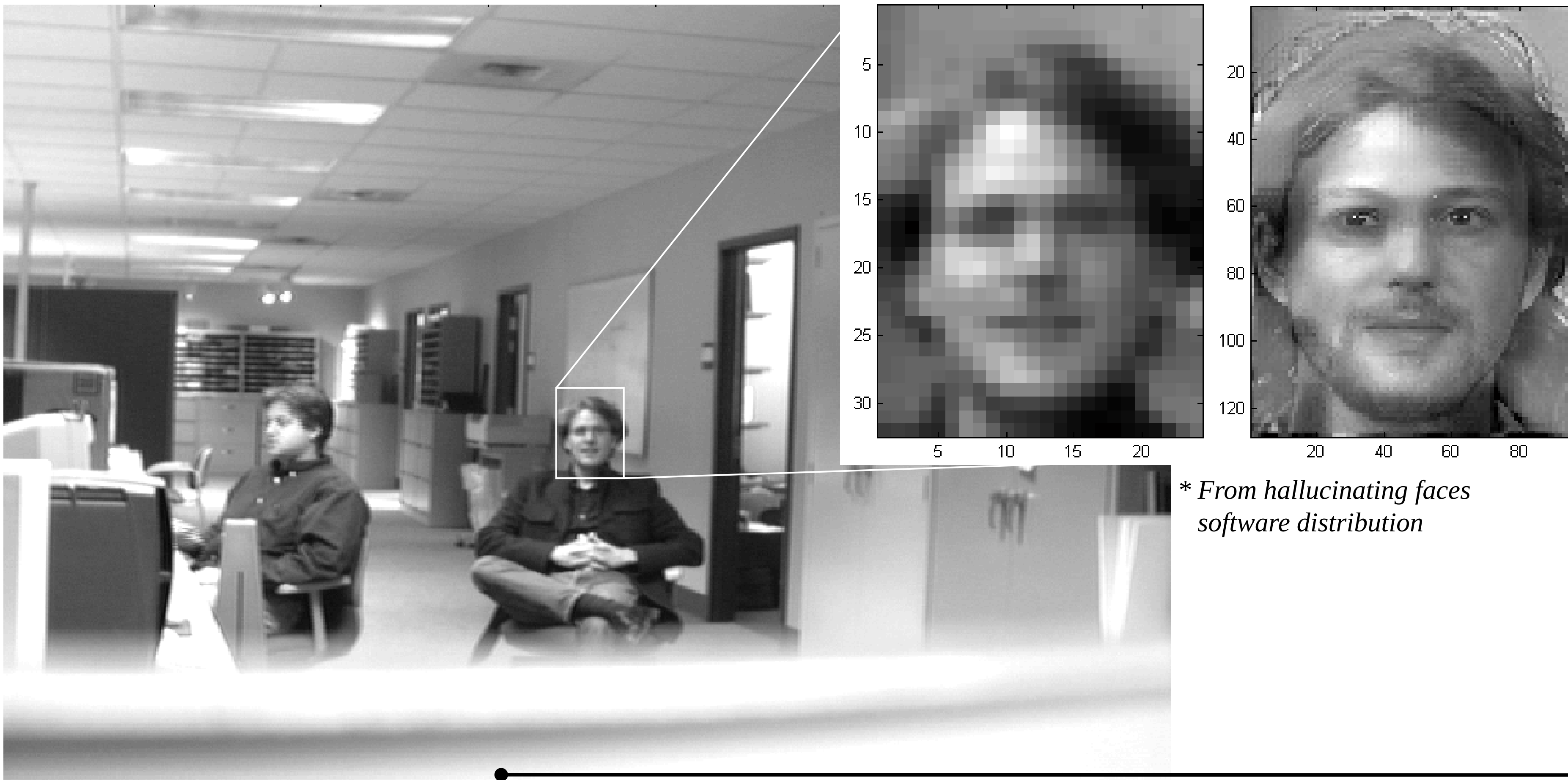


Reconstruction of Low-resolution Images

- We could try to reconstruct the face image
- Reconstruction algorithms find estimates of missing HR pixels in a LR image by assuming a model



Super-Resolution with Face Priors



** From hallucinating faces software distribution*

S. Baker and T. Kanade, "Limits on Super-Resolution and How to Break Them", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 24, No. 9, September, 2002, pp. 1167 - 1183.

The Image Formation Process

- Inverting the image formation process

$$y = \mathbf{DHW}x + n$$

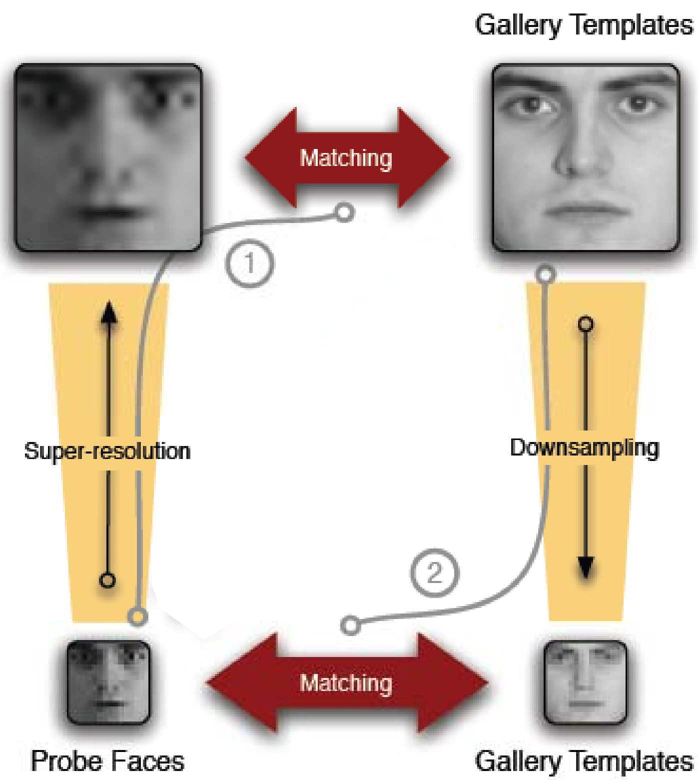
- y: low-resolution image
- $\mathbf{D}$: decimation matrix
- $\mathbf{H}$: blur matrix
- $\mathbf{W}$: warp matrix
- x: high-resolution image
- n: detector noise

- Super-resolution algorithms aim to invert this process either directly or indirectly.

See, for example:

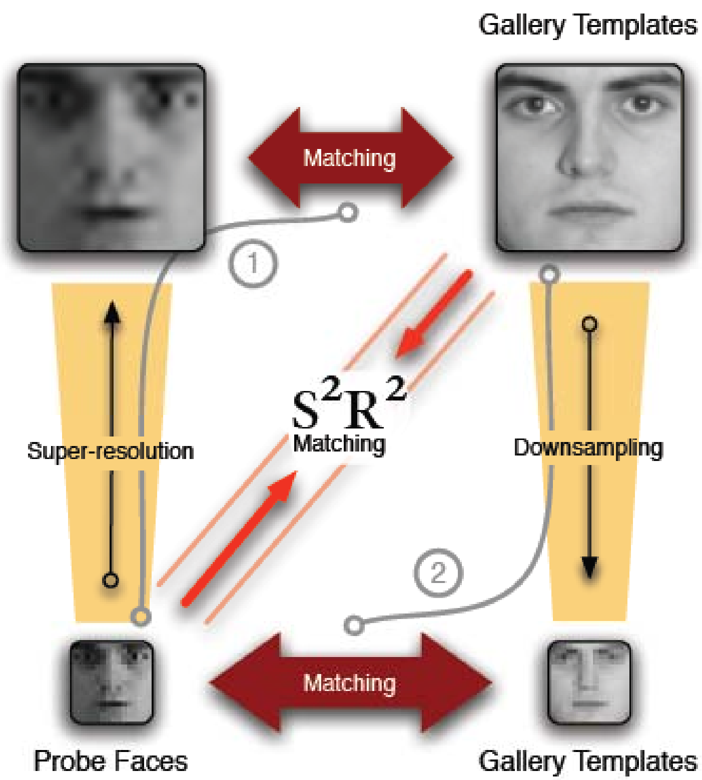
A. Zomet and S. Peleg, "Super-Resolution from Multiple Images having Arbitrary Mutual Motion", in S. Chaudhuri (Editor), Super-Resolution Imaging, Kluwer Academic, Sept. 2001, pp. 195-209.

Possible Solutions



1. We can apply a super-resolution algorithm and then classify the result
2. We can downsample the gallery image and match at the resolution of the probe

Proposed Solution: S^2R^2

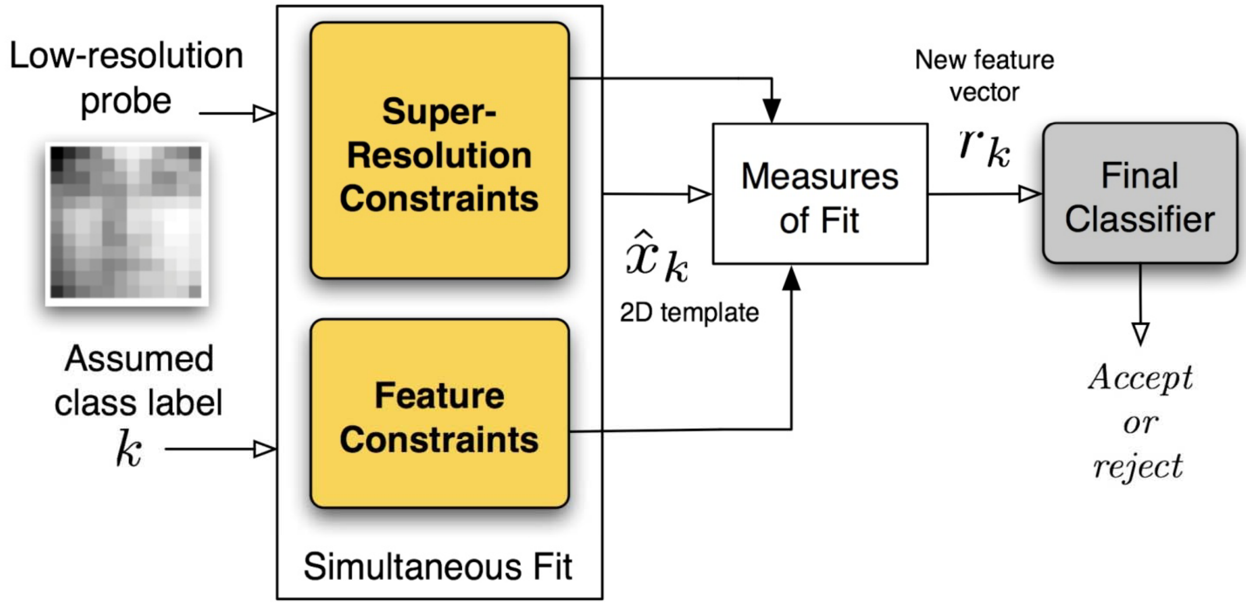


1. We can apply a super-resolution algorithm and then classify the result

2. We can downsample the gallery image and match at the resolution of the probe

- We propose an alternative approach which jointly uses super-resolution methods and includes face features for recognition (S^2R^2)

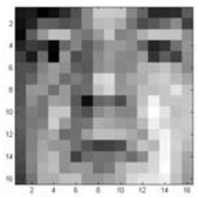
S^2R^2 Block Diagram



S^2R^2 Simultaneous fit

- Minimize the regularized functional given by

$$J(x;k) = \|\mathbf{B}x -$$



$$\|^2 + \alpha^2 \|\mathbf{L}x\|^2 + \beta^2 \|\mathbf{F}x - f^{(k)}\|^2$$

Index of class to match (k)

Image formation matrix ($\mathbf{B}$)

Low-res **probe**

First-derivatives matrix ($\mathbf{L}$)

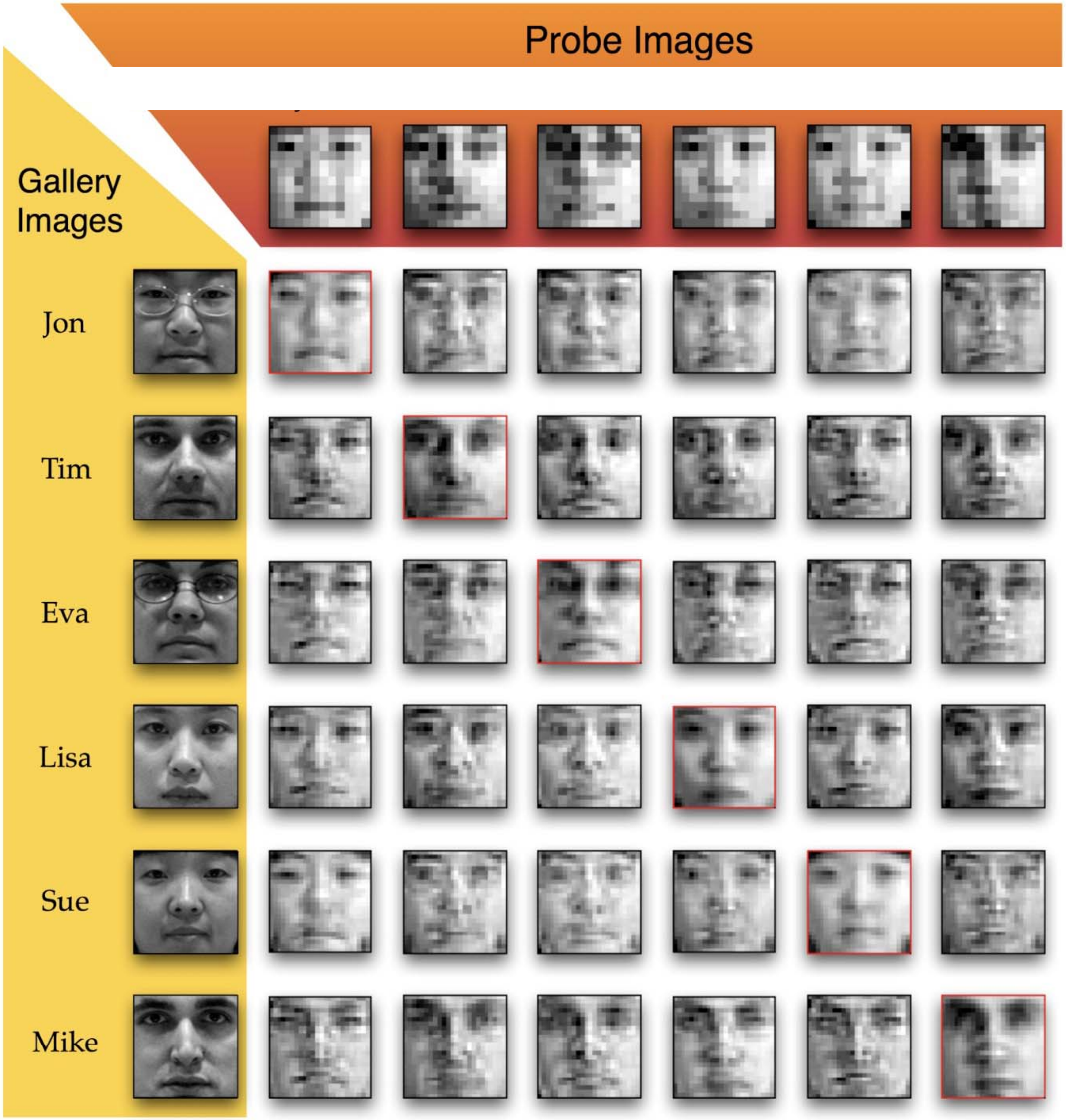
Feature extraction matrix ($\mathbf{F}$)

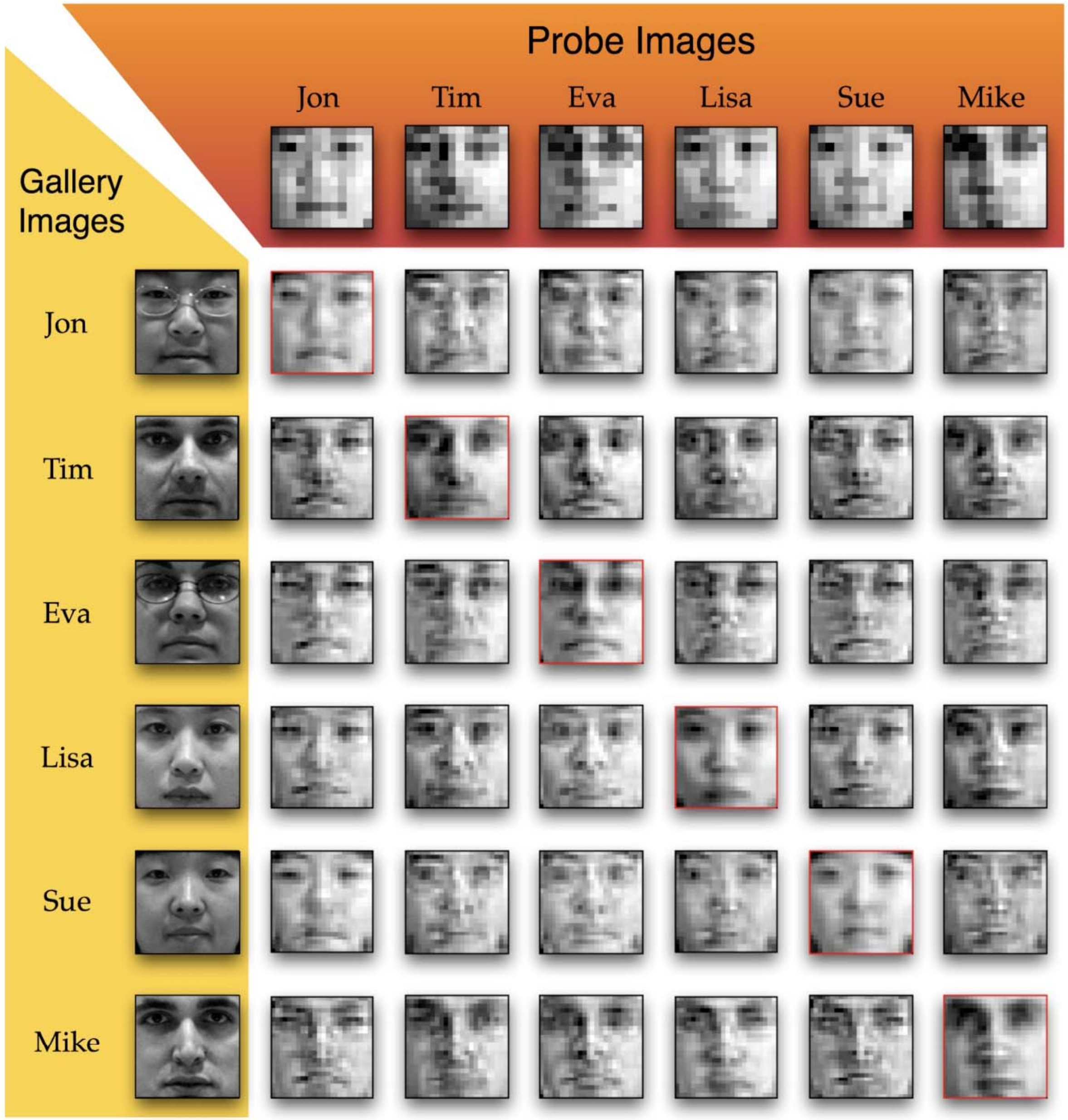
Features from **gallery** image ($f^{(k)}$)

Super-resolution models

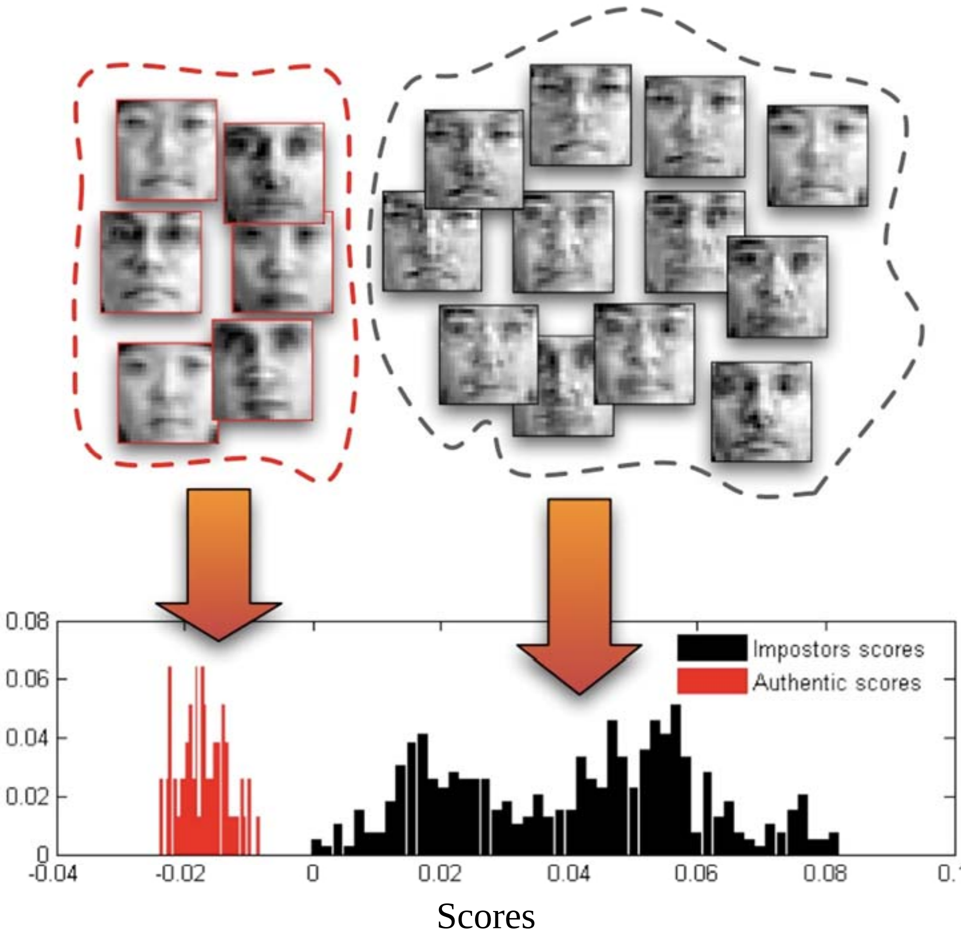
Feature extraction model

- Regularization parameters are trained to produce distortions that are discriminatory


Probe Images
Gallery Images
Jon
Tim
Eva
Lisa
Sue
Mike


Probe Images
Jon
Tim
Eva
Lisa
Sue
Mike
Gallery Images
Jon
Tim
Eva
Lisa
Sue
Mike

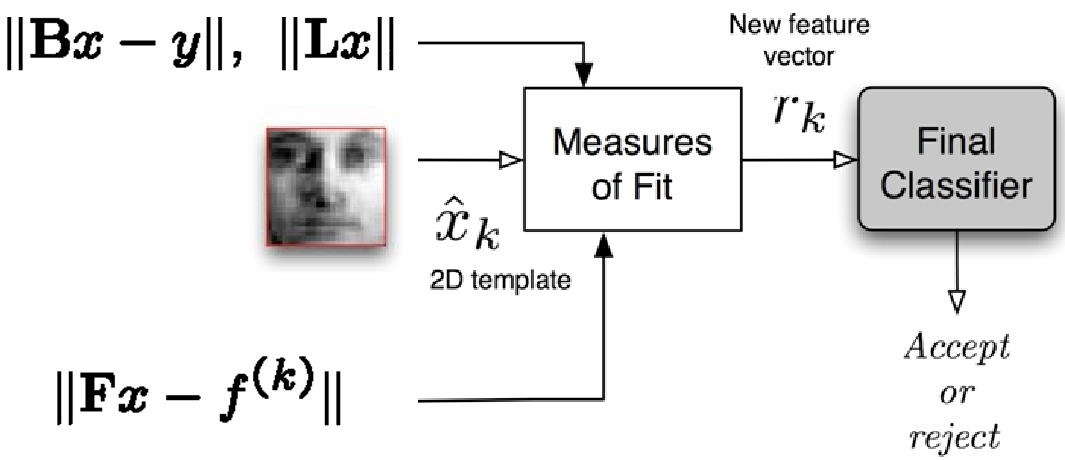
S^2R^2 Classification Main Idea



- By computing measures of fit from these simultaneous-fit results we extract new features.
- A binary classifier is trained such that the scores of true-fits and false-fits using the new features are as separate as possible.
- Here, a linear discriminant is used.

S^2R^2 Classification

- Compute measures-of-fit norms and form a new feature vector r_k

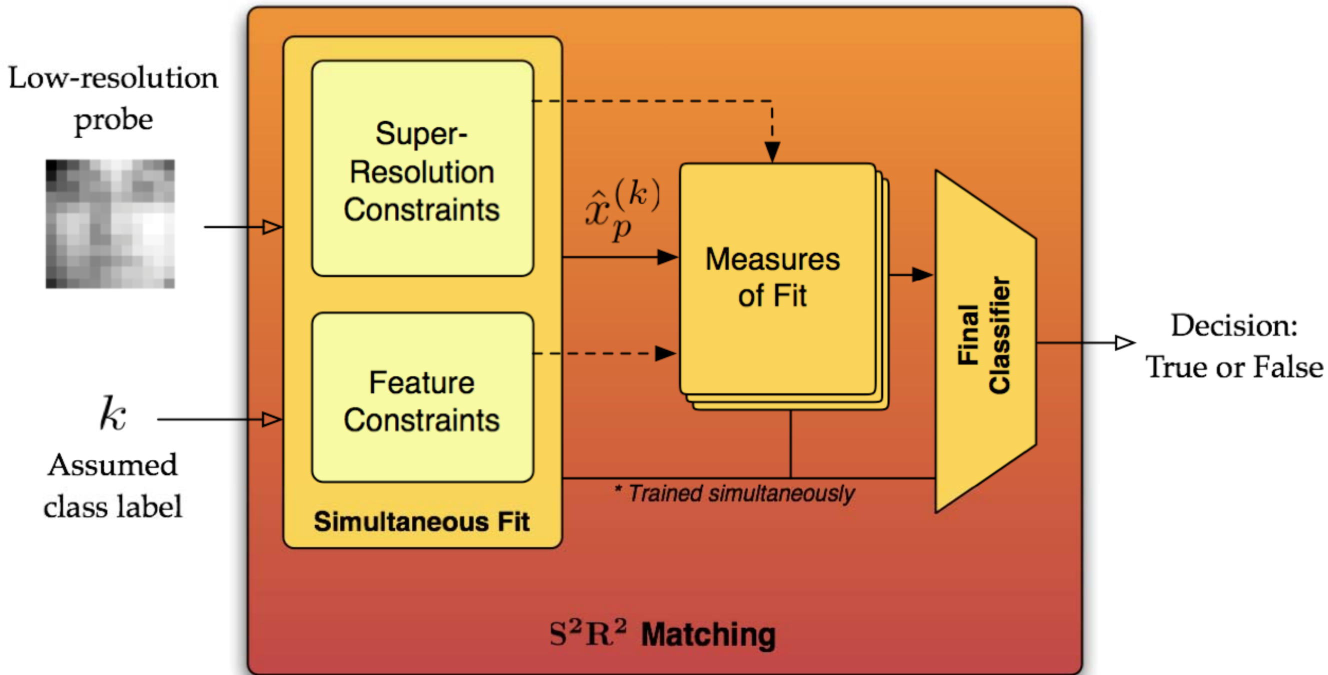


- Classify with r_k using conventional classification methods

$$r_k = \begin{bmatrix} \|\mathbf{B}\hat{x}_k - y\| \\ \|\mathbf{L}\hat{x}_k\| \\ \|\mathbf{F}\hat{x}_k - f^{(k)}\| \end{bmatrix}$$

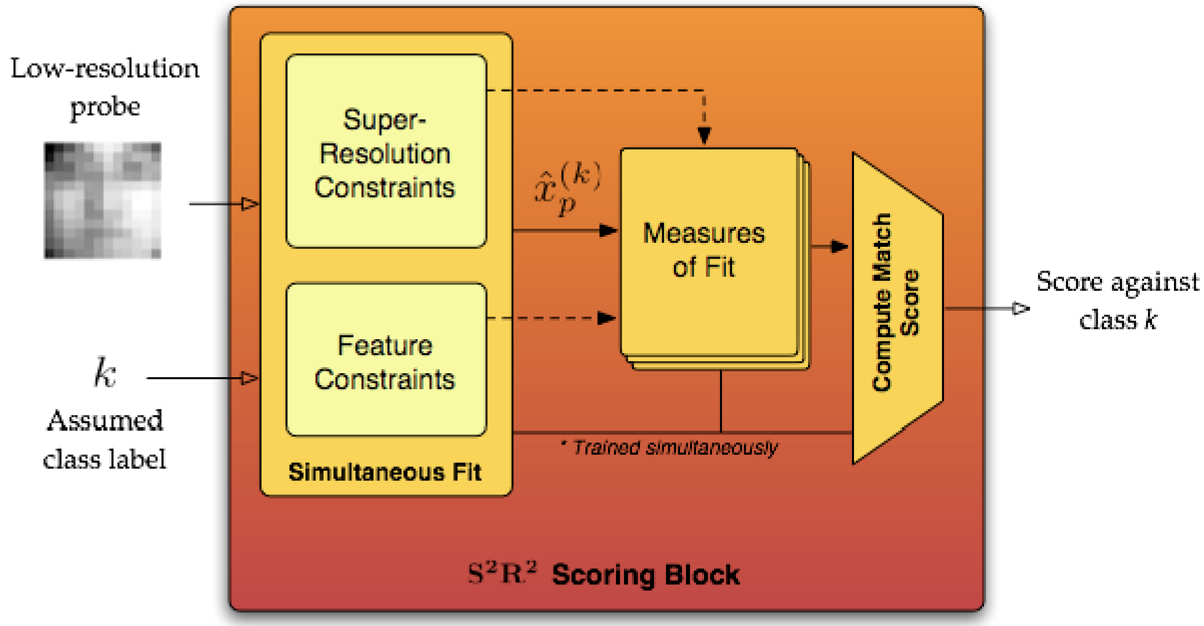
Verification with S^2R^2

- In verification a claimed class is input with the LR probe



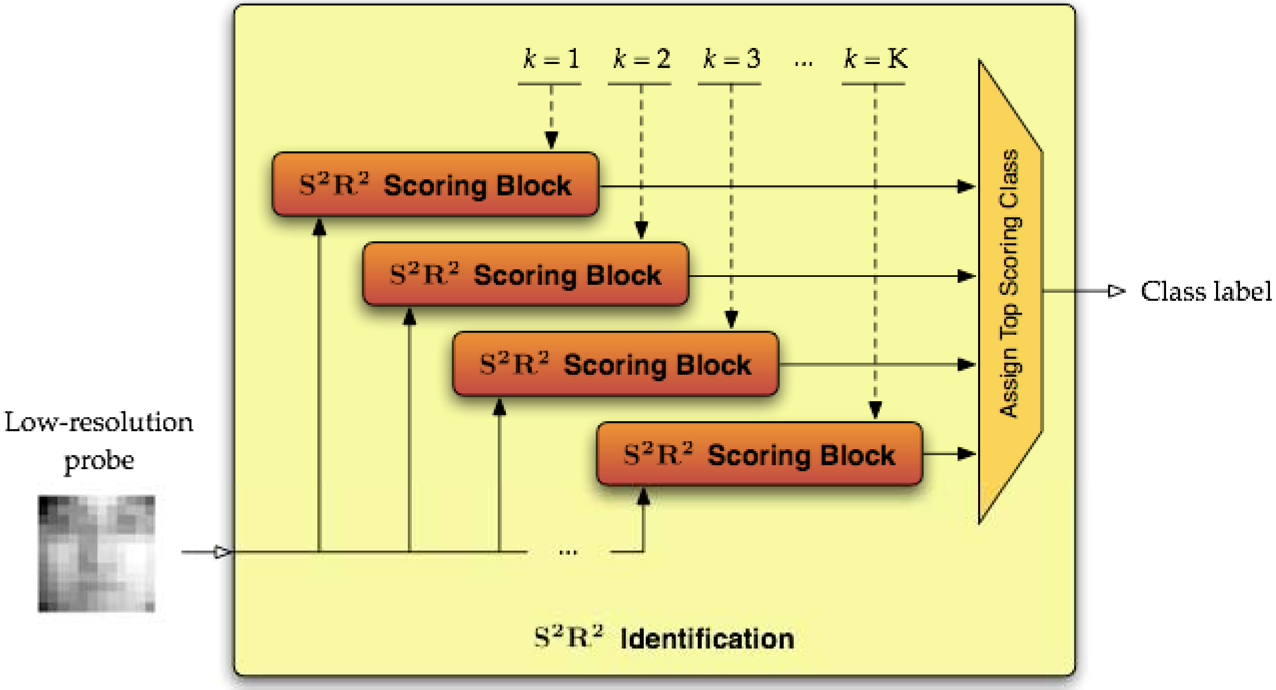
Scoring with S²R²

- In general, a classifier can output a score instead of a hard decision



Identification with S^2R^2

- In identification the input probe is scored against K classes



S^2R^2 Implementation

- The simultaneous fit of S^2R^2 is an LS problem:

$$J(x;k) = \|\mathbf{B}x - y\|^2 + \alpha^2\|\mathbf{L}x\|^2 + \beta^2\|\mathbf{F}x - f^{(k)}\|^2$$

- Typically solved using iterative methods

- But for every LR input image y, and for every claimed class k we have to solve a simultaneous-fit and extract measures-of-fit features

$$r_k = \begin{bmatrix} R_B \\ R_L \\ R_F \end{bmatrix} = \begin{bmatrix} \|\mathbf{B}\hat{x} - y\| \\ \|\mathbf{L}\hat{x}\| \\ \|\mathbf{F}\hat{x} - f^{(k)}\| \end{bmatrix}$$

We need to do thousands or millions of comparisons.

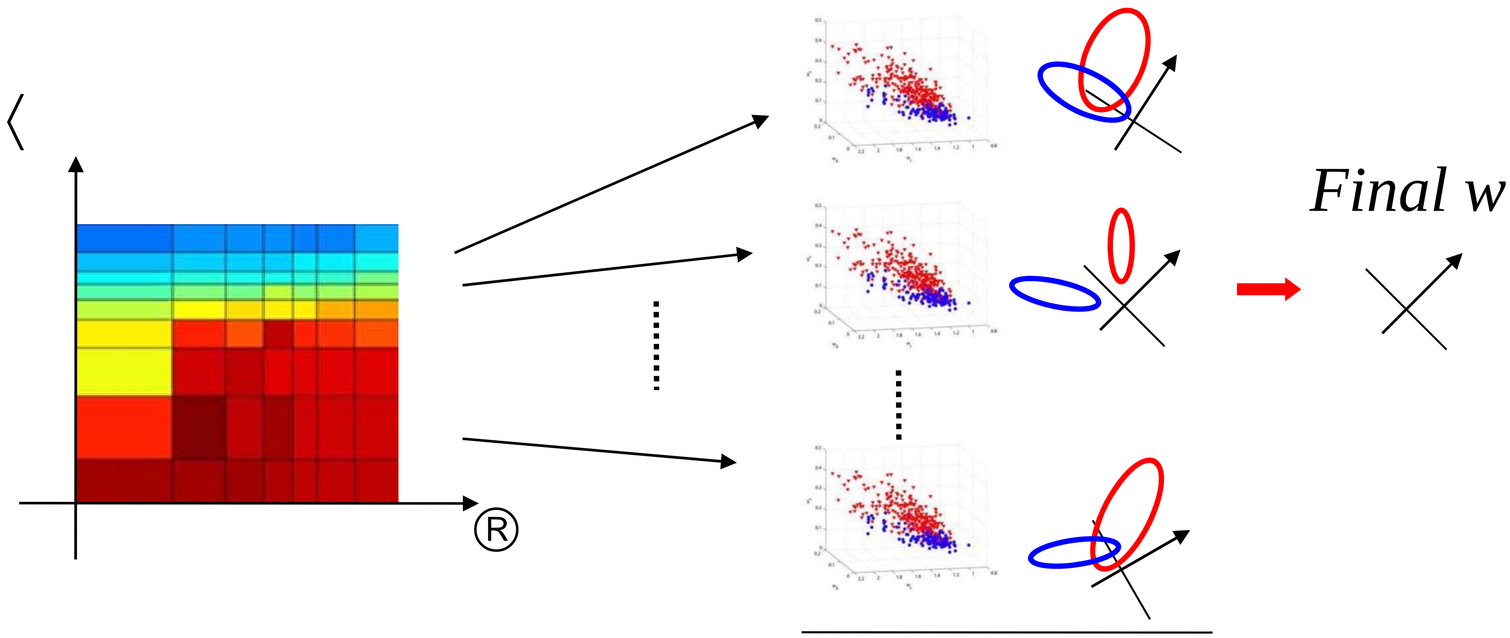
Simultaneous Fit Using SVD methods

- Let
$$\mathbf{G} = \begin{bmatrix} \mathbf{B} \\ \alpha\mathbf{L} \\ \beta\mathbf{F} \end{bmatrix} = \mathbf{U}\mathbf{S}\mathbf{V}^T \qquad \mathbf{G}^\dagger = \mathbf{V}\mathbf{S}^{-1}\mathbf{U}^T$$
- Then we can write the reconstructed image as
$$\hat{x}_k = \mathbf{V}\mathbf{S}^{-1}\mathbf{U}^H d = \mathbf{G}^\dagger d$$
$$\hat{x}_k = \mathbf{G}_1^\dagger d_1 + \mathbf{G}_2^\dagger d_2 + \mathbf{G}_3^\dagger d_3$$
$$\begin{aligned} d_1 &= y \\ d_2 &= 0 \\ d_3 &= \beta f^{(k)} \end{aligned}$$
- This becomes
$$\hat{x}_k = \mathbf{G}_1^\dagger y + \beta \mathbf{G}_3^\dagger f^{(k)} = \mathbf{G}_1^\dagger y + m^{(k)}$$

Can be pre-computed before testing this probe image against each classifier

Finding Regularization Parameters and Final Classifier

- In training, all the regularization parameters are found simultaneously with the final classifier *w*



At every step, compute error metric of classifier to rank each (〈, ®)

S^2R^2 is a General Framework

- It consists of a simultaneous fit formulation

$$\hat{x}_p^{(k)} = \underset{x}{\operatorname{argmin}} \|\mathbf{B}x - y_p\|^2 + \alpha^2\|\mathbf{L}x\|^2 + \beta^2\|\mathbf{F}x - f_g^{(k)}\|^2$$

- And measures-of-fit features and classifier

$$q_{\hat{x}_p^{(k)}} = \begin{bmatrix} \|\mathbf{B}\hat{x}_p^{(k)} - y_p\|^2 \\ \|\mathbf{L}\hat{x}_p^{(k)}\|^2 \\ \|\mathbf{F}\hat{x}_p^{(k)} - f_g^{(k)}\|^2 \end{bmatrix}$$

Verification

$$\omega \leftarrow \left[w \cdot q_{\hat{x}_p^{(k)}} \right]_t$$

Identification

$$\omega \leftarrow \underset{k}{\operatorname{argmin}}\, w \cdot q_{\hat{x}_p^{(k)}}$$

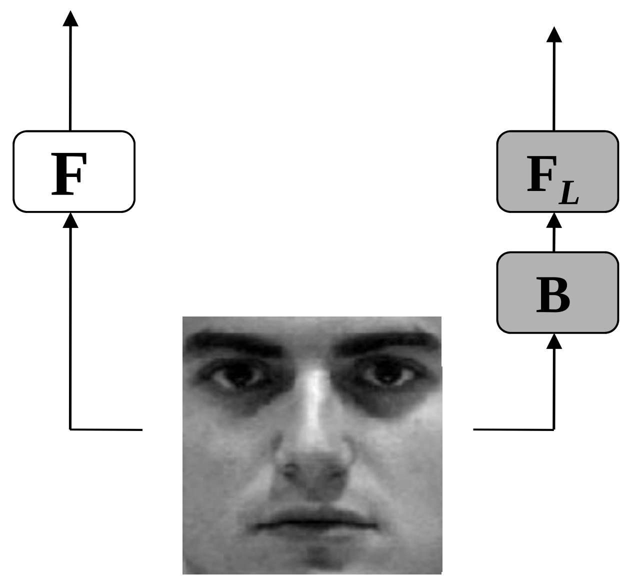
Using Features at Multiple Resolutions (MFS^2R^2)

- Since we know the image formation process

$$\underset{x}{\operatorname{argmin}} \|\mathbf{B}x - y_p\|^2 + \alpha^2\|\mathbf{L}x\|^2 + \beta^2\|\mathbf{F}x - f^{(k)}\|^2 + \gamma^2\|\mathbf{F}_L\mathbf{B}x - f_L^{(k)}\|^2$$

- Features defined as

$$q_{\hat{x}_p^{(k)}} = \begin{bmatrix} \|\mathbf{B}\hat{x}_p^{(k)} - y_p\|^2 \\ \|\mathbf{L}\hat{x}_p^{(k)}\|^2 \\ \|\mathbf{F}\hat{x}_p^{(k)} - f^{(k)}\|^2 \\ \|\mathbf{F}_L\mathbf{B}\hat{x}_p^{(k)} - f_L^{(k)}\|^2 \end{bmatrix}$$

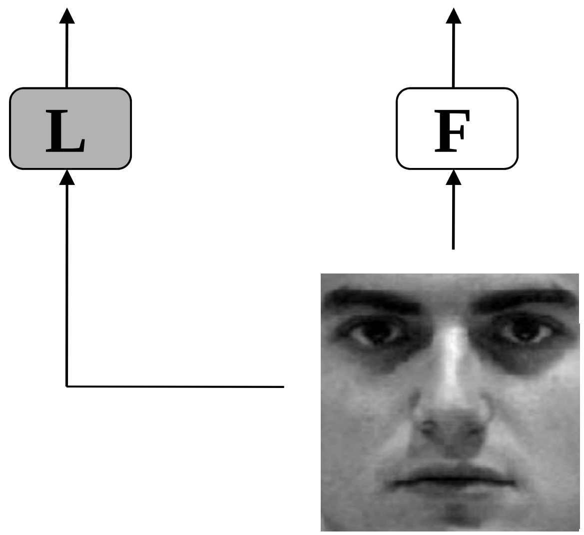


Gallery image form class *k*

Simultaneous Fits Using Face Priors

- In training, we can compute super-resolution priors for every gallery image

$$\hat{x}_p^{(k)} = \underset{x}{\operatorname{argmin}} \|\mathbf{B}x - y_p\|^2 + \alpha^2 \|\mathbf{L}x - l_g^{(k)}\|^2 + \beta^2 \|\mathbf{F}x - f_g^{(k)}\|^2$$



Gallery image form class *k*

S^2R^2 Features Using Relative Residuals

- Using a simultaneous fit given by

$$\hat{x}_p^{(k)} = \underset{x}{\operatorname{argmin}} \|\mathbf{B}x - y_p\|^2 + \alpha^2 \|\mathbf{L}x - l_g^{(k)}\|^2 + \beta^2 \|\mathbf{F}x - f_g^{(k)}\|^2$$

- Normalization of features helps the final classification stage

$$q_{\hat{x}_p^{(k)}} = \begin{bmatrix} \|\mathbf{B}\hat{x}_p^{(k)} - y_p\|^2 / \|y_p\|^2 \\ \|\mathbf{L}\hat{x}_p^{(k)} - l_g^{(k)}\|^2 / \|l_g^{(k)}\|^2 \\ \|\mathbf{F}\hat{x}_p^{(k)} - f_g^{(k)}\|^2 / \|f_g^{(k)}\|^2 \end{bmatrix}$$

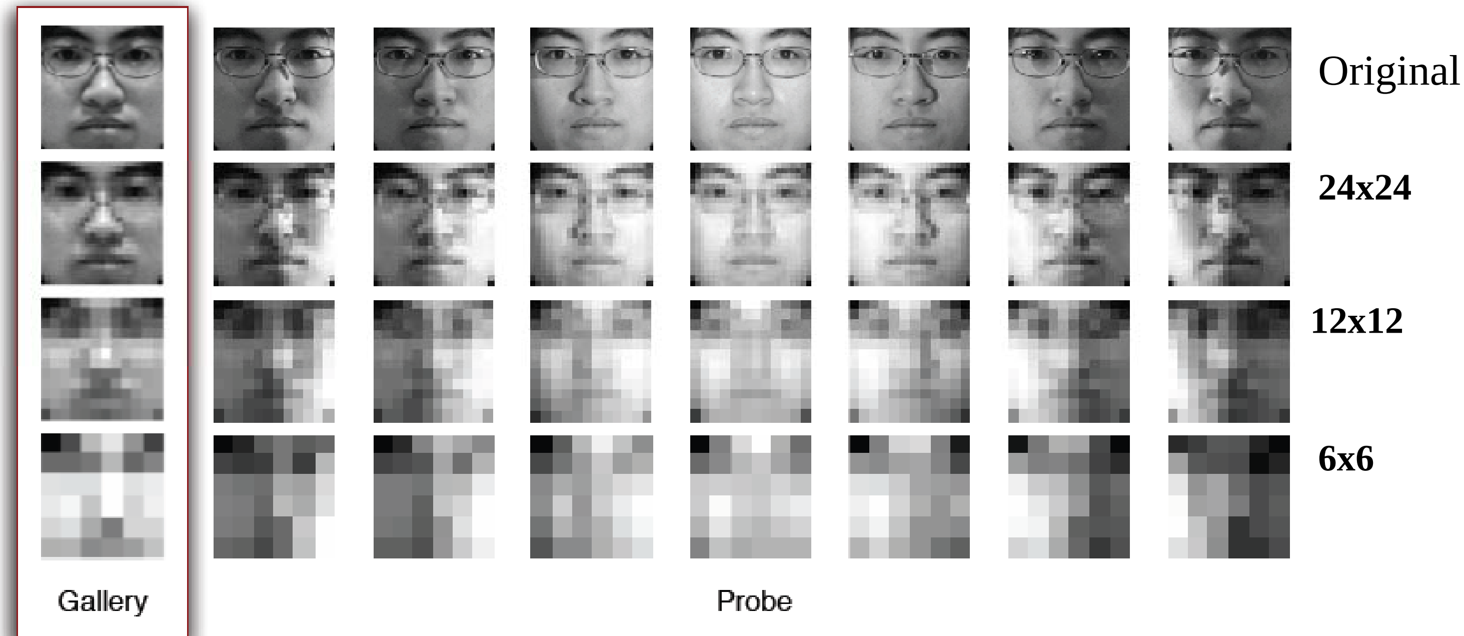
- We refer to this version of the proposed framework as S^2R^2e

Database

- The MultiPIE database
  – Total of 337 subjects, compared to 68 of PIE
  – Subjects are captured in several recording sessions with different poses, illuminations and expressions, as in PIE

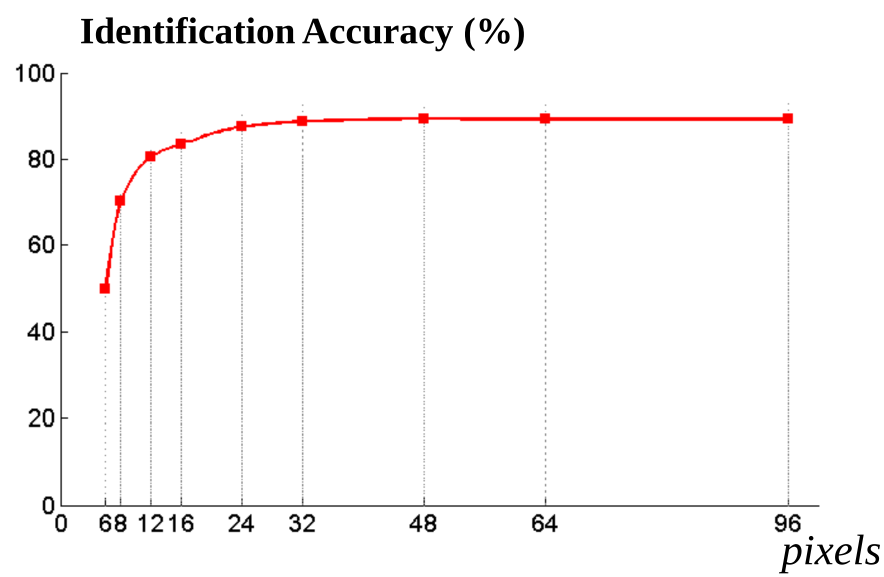
- Data set for experiments
  – Using frontal view, neutral expressions, different flash illuminations
  – Sequestered 73 subjects as generic training set
  – Sequestered 40 subjects to learn regularization parameters
  – **The rest 224 subjects are used as gallery and probes**
  – Gallery images are not under flash illumination
  – **We use only 1 image per gallery class**
  – There are a total of 2912 probe images

Sample from Multi-PIE



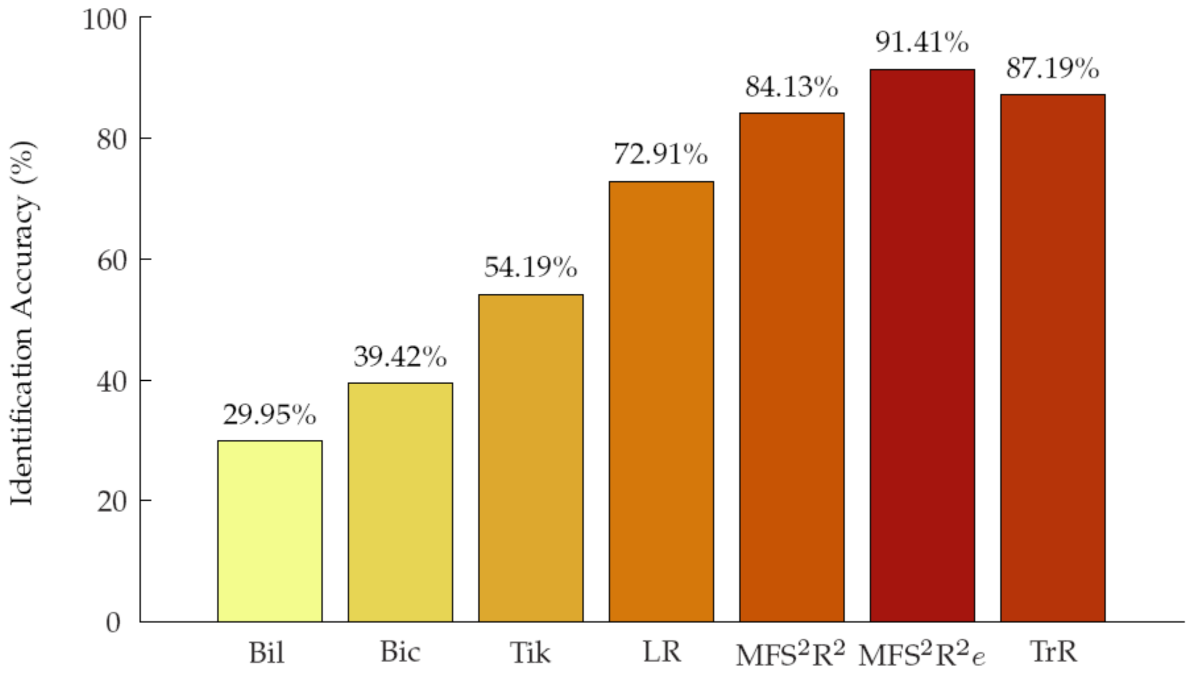
Experimental Setup

- Proposed framework settings
  - Using a single input (still image) probe
  - The image formation process is assumed known
  - Using smoothness constraints as first derivatives
  - Face-features are 25 Fisherfaces
  - Final classifier is a Fisher discriminant
  - Training resolution is 24x24 pixels
  - S^2R^2 using face-features at training resolution and probe resolution.



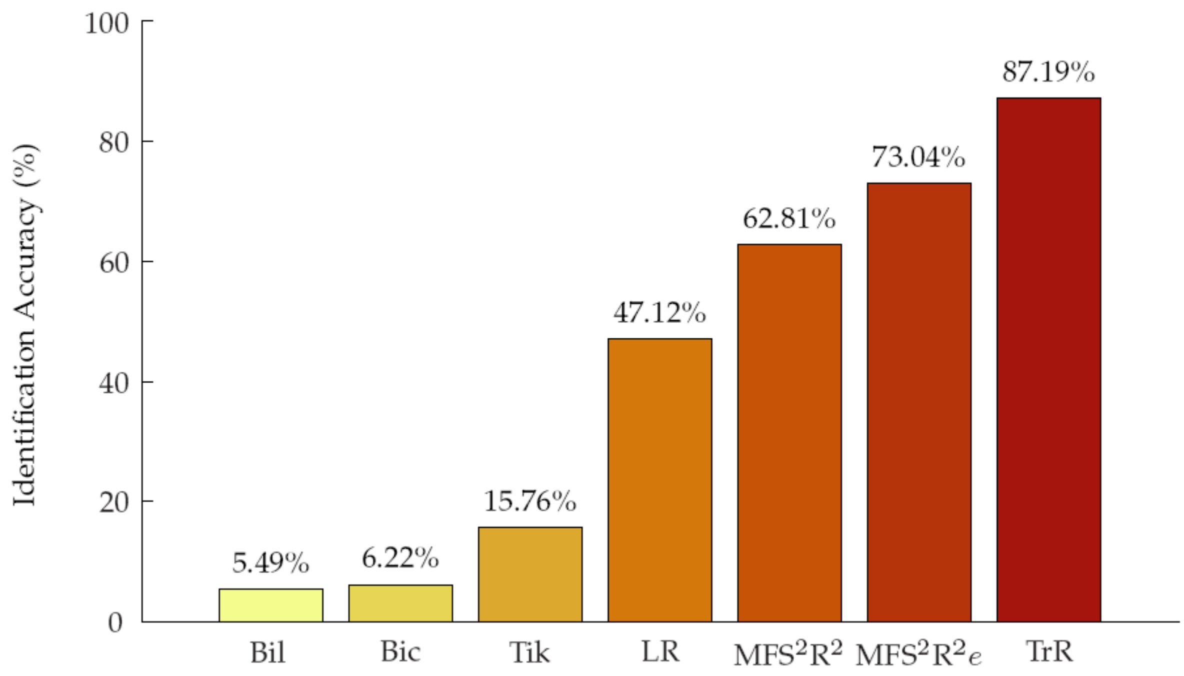
MFS^{2}R^2*e* Results Using Multi-PIE (I)

- Magnification factor of 2



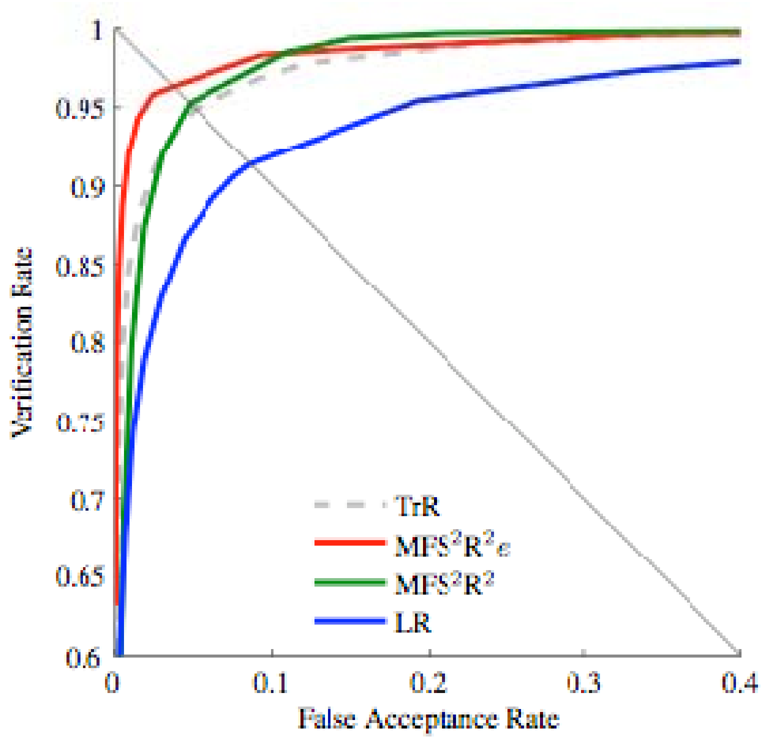
MFS^{2}R^2*e* Results Using Multi-PIE (II)

- Magnification factor of 4

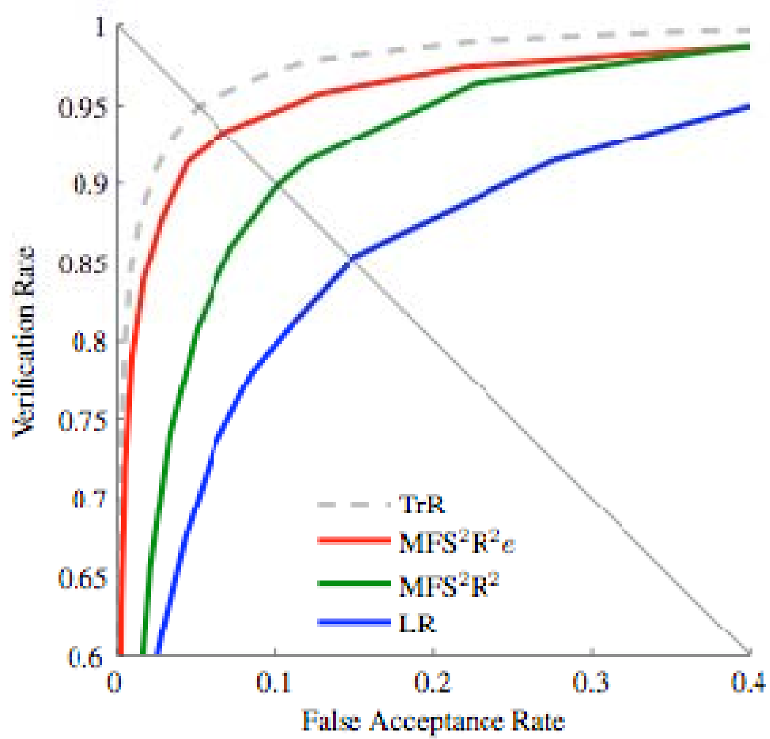


ROC Curves

Magnification factor of 2



Magnification factor of 4



Conclusions

- Before our work, there were two conventional approaches for recognition of low-resolution faces, both with clear disadvantages.
- Most of the effort in related works focuses on image reconstruction using super-resolution, with face recognition addressed only as an afterthought.
- Now, with the proposed S^2R^2 framework, super-resolution methods can be tailored to face recognition, rather than just reconstruction.
- The proposed framework can use existing super-resolution and feature extraction models. It extracts new features by finding a template that fits simultaneously into the available models and face-features.
- We have shown that with simple linear discriminants using these new (measures-of-fit) features we can produce better recognition performance than conventional approaches.
- Finally, the proposed framework is flexible. Our formulation can be easily expanded or generalized to use video, multiple cameras, and even other image representations (such as wavelets) and non-linear features.

Thank you