3D-Assisted Facial Texture Super-Resolution

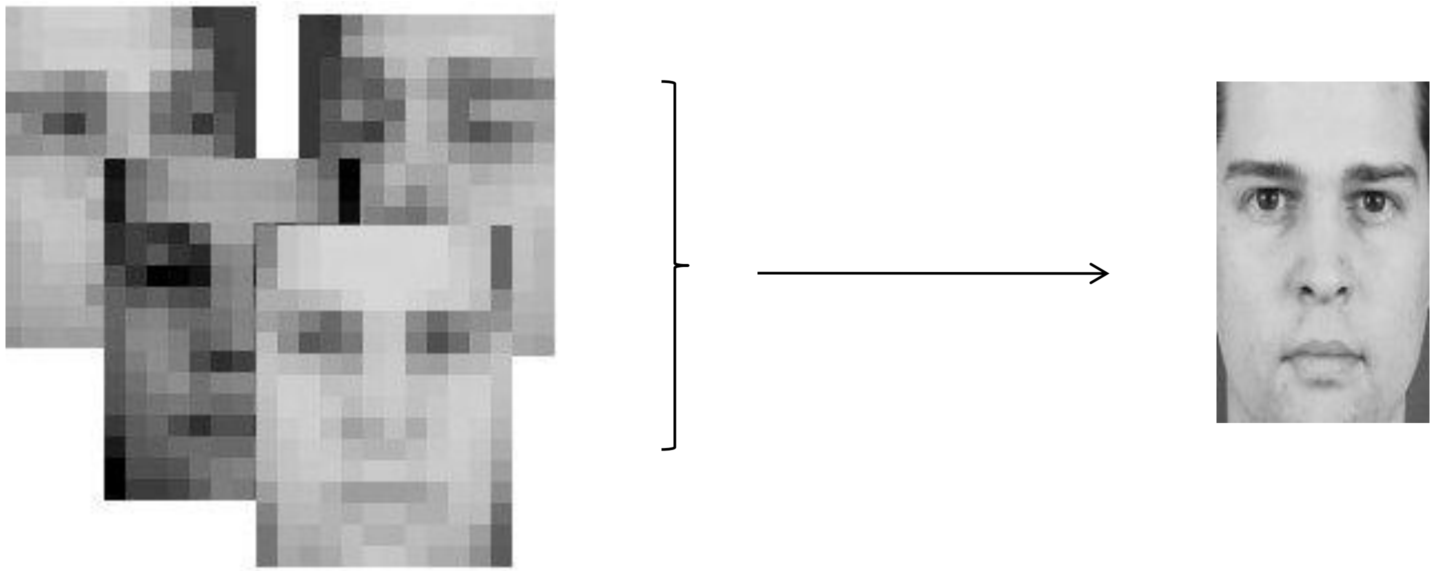
Pouria Mortazavian, Josef Kittler, William Christmas

Centre for Vision, Speech and Signal Processing
University of Surrey
p.mortazavian@surrey.ac.uk

10 September 2009

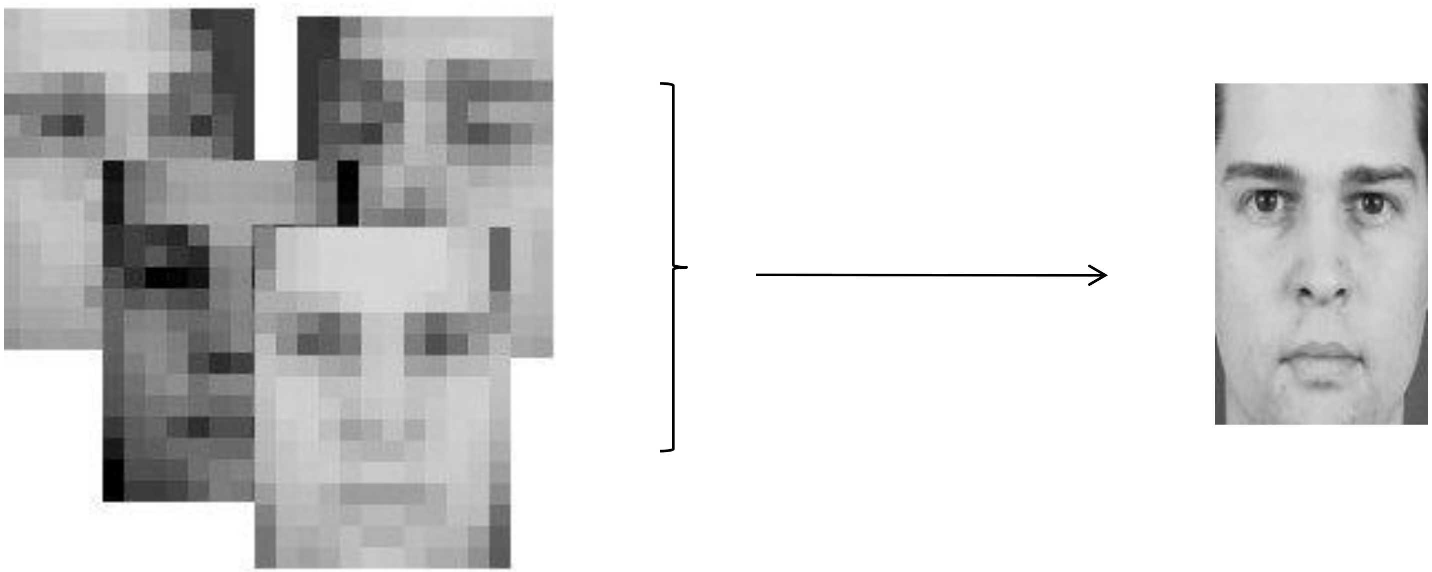
Super-Resolution:

Given a number of low-resolution observations from the same scene/object, estimate a high resolution image of that scene/object.



Super-Resolution:

Given a number of low-resolution observations from the same scene/object, estimate a high resolution image of that scene/object.



- Reconstruction-based
- Example-based
  - Object-specific

Maximum aposteriori estimation:

$$F^* = \arg\max_F \{ \prod_i p(F \mid f_i) \} = \arg\max_F \{ \prod_i p(f_i \mid F) p(F) \}$$

Maximum aposteriori estimation:

$$F^* = \arg\max_F \{ \prod_i p(F \mid f_i) \} = \arg\max_F \{ \prod_i p(f_i \mid F)\, p(F) \}$$

Likelihood:

<u>SR constraint</u>: The HR image, when appropriately warped and down-sampled should yield the LR input images.

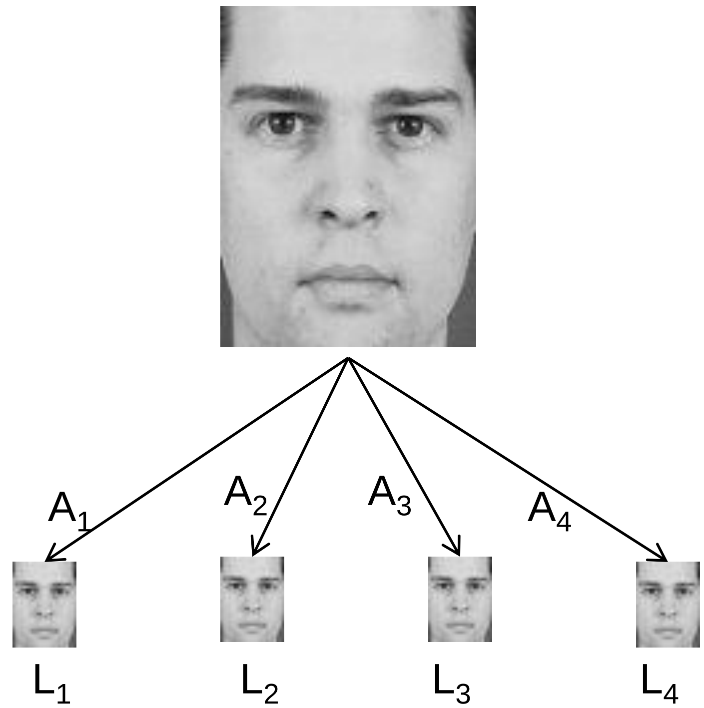
$$f_i = \mathbf{A_i} \,.\, F + \eta_i$$

A : Warp, Blur, Down Sampling

η : Pixel noise

$$- \log P(\boldsymbol{f}_i | \boldsymbol{F}) \sim \sum_i \| A_i \boldsymbol{F} - \boldsymbol{f}_i \|^2$$

Generative Model :



Face Hallucination:
[Baker and Kanade, PAMI 2002]

- Likelihood

$$- \log P(\boldsymbol{f}_i | \boldsymbol{F}) \sim \sum_i || A_i \boldsymbol{F} - \boldsymbol{f}_i ||^2$$

- Prior:
  - Gradient Prediction

$$- \log p(\boldsymbol{F}) \sim \sum_{m,n,i} [\widehat{\nabla \boldsymbol{F}}_{\mathrm{i}}(\mathrm{m,n}) - \nabla \boldsymbol{F}_{\mathrm{i}}(\mathrm{m,n})]^2$$

Face Hallucination:
[Baker and Kanade, PAMI 2002]

3D Morphable Model:

• A ***3D Morphable face model*** represents each face by a set of model coefficients, and generates new, natural-looking faces from any novel set of coefficients.

• 3D structure of a known face is captured in *shape* and *texture* vectors

$$S_{model} = \overline{S} + \sum_{i=1}^{m} \alpha_i S_i$$

$$R_{model} = \overline{R} + \sum_{i=1}^{m} \beta_i R_i$$

Fitting the 3DMM to 2D Images:

• Model parameters (α, β, ρ) are optimized using a MAP estimator such that the appearance of the model matches that of the 2D image.

$$\{\alpha^*, \beta^*, \rho^*\} = \arg\max_{\alpha,\beta,\rho} p(\alpha,\beta,\rho \mid F) = p(F \mid \alpha,\beta,\rho).p(\alpha,\beta,\rho)$$

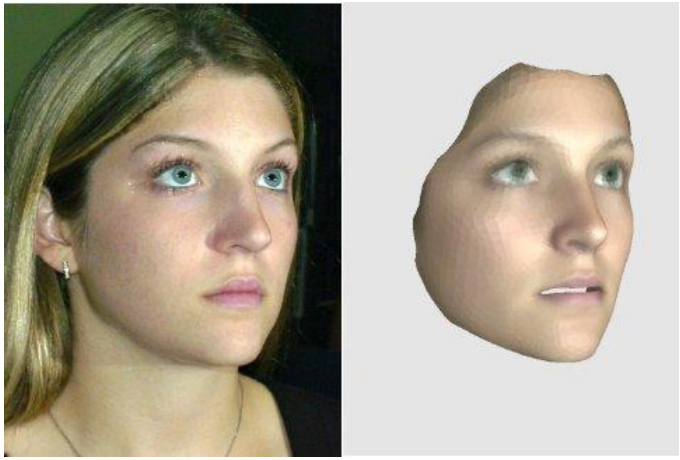


Image taken from J.R. Tena Rodr´ıguez's PhD thesis: "*3D Face Modelling for 2D+3D Face Recognition*"

Texture Extraction:

• Once the model is fitted on a 2D image, we can extract the texture from the input image and map to a pre-defined, shape- and pose-normalized coordinate frame:

image & model projection   optimised model   texture map

Image taken from J.R. Tena Rodr´ıguez's PhD thesis: "*3D Face Modelling for 2D+3D Face Recognition*"

3D-Assisted SR:

$$T^* = \arg\max_T \sum_{\mu,\rho} p(T, \mu, \rho \mid f)$$

$$= \arg\max_T \sum_{\mu,\rho} p(T \mid \mu, \rho, f).p(\mu, \rho \mid f)$$

3D-Assisted SR:

UNIVERSITY OF SURREY

$$T^* = \arg\max_T \sum_{\mu,\rho} p(T,\mu,\rho \mid f)$$

$$= \arg\max_T \sum_{\mu,\rho} \underbrace{p(T \mid \mu,\rho,f)}_{\textit{Texture SR}} . \underbrace{p(\mu,\rho \mid f)}_{\textit{Model Fitting}}$$

3D-Assisted SR:

$$T^* = \arg\max_T \sum_{\mu,\rho} p(T,\mu,\rho \mid f)$$

$$= \arg\max_T \sum_{\mu,\rho} \underbrace{p(T \mid \mu,\rho,f)}_{\text{Texture SR}} . \underbrace{p(\mu,\rho \mid f)}_{\text{Model Fitting}}$$

Assuming μ and ρ have a dense distribution which peaks at their optimal value (obtained by model fitting), the above simplifies to:

$$T^* = \arg\max_T \; p(T \mid \mu^*, \rho^*, f)$$

3D-Assisted SR:

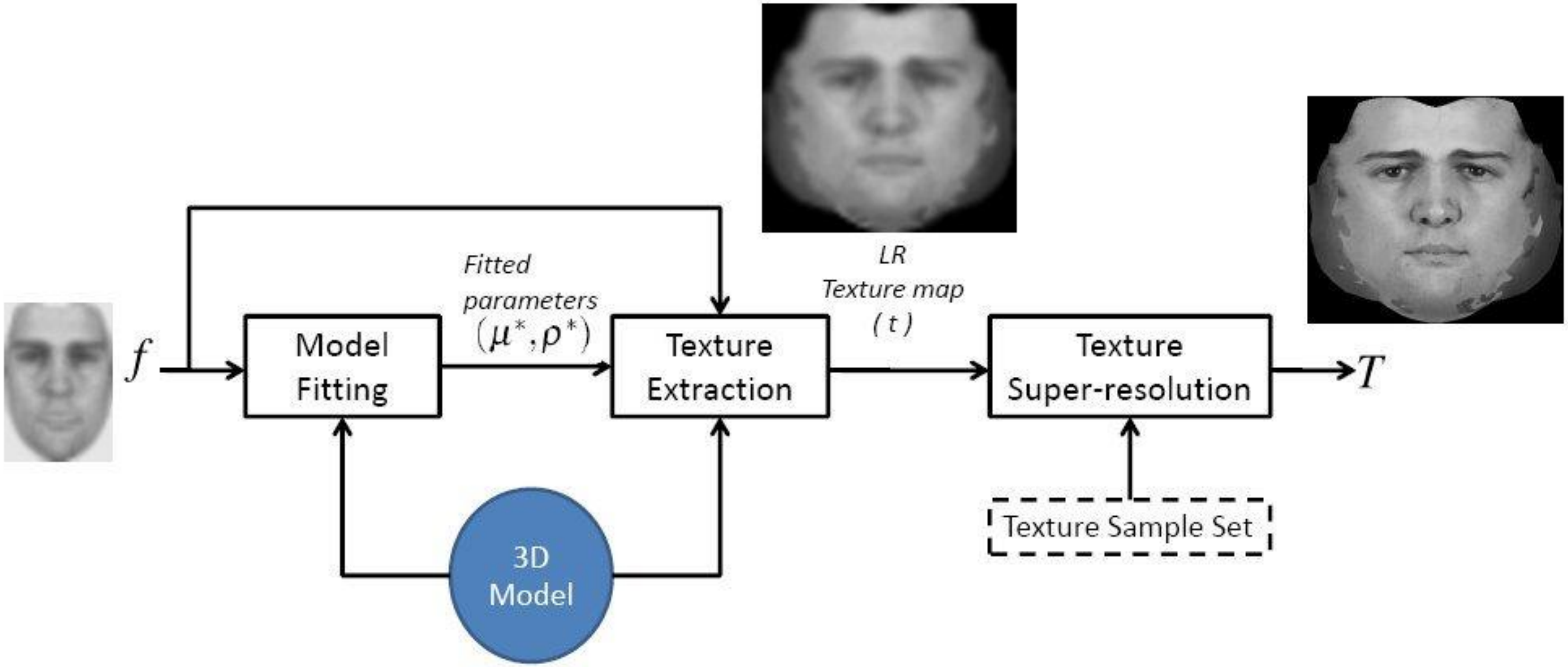
let:

$$t = \mathrm{TEXTURE\_EXTRACT}(\mu^*, \rho^*, f)$$

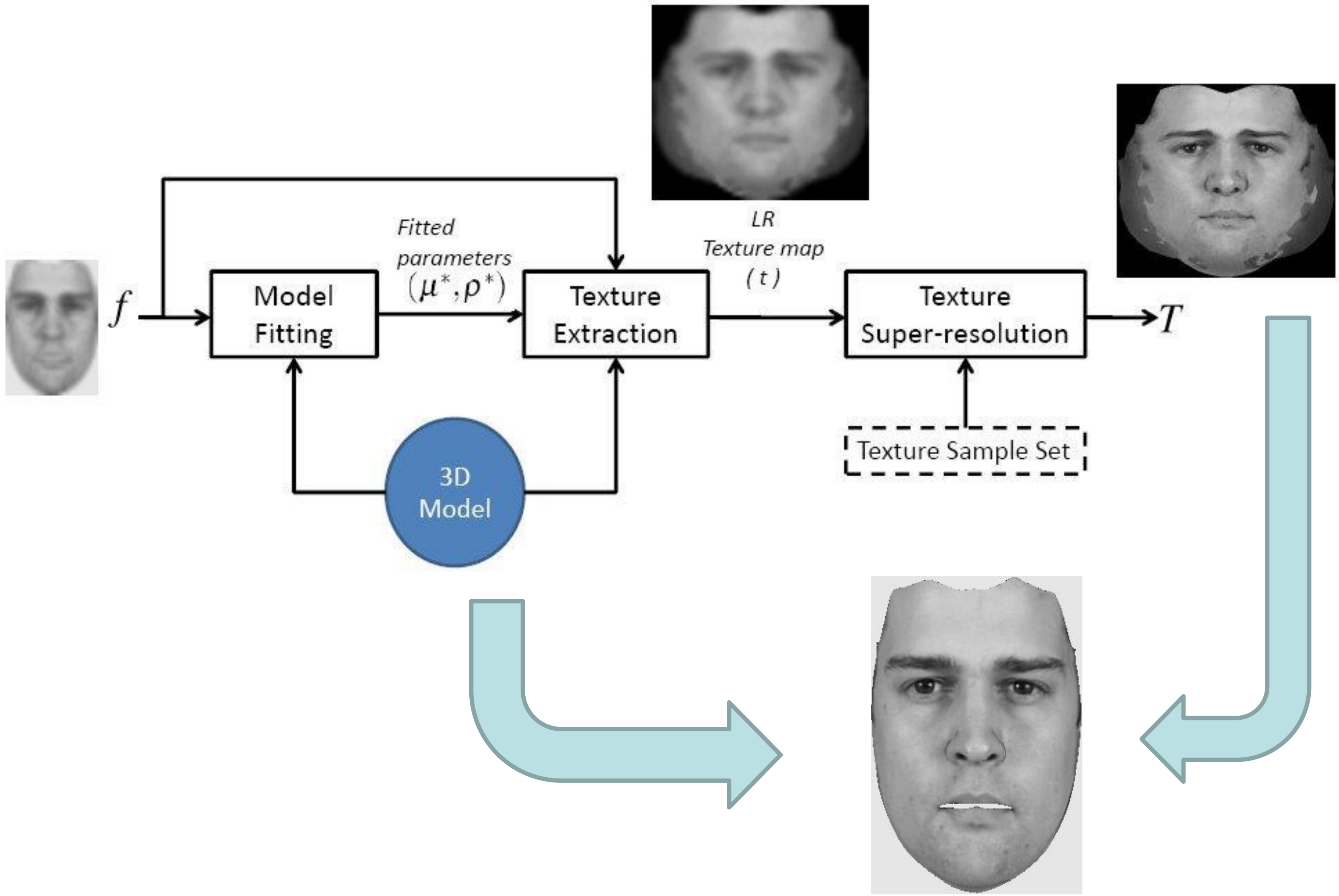
Assuming t has all information available in f:

$$T^* = \arg\max_T \; p(T \mid t) = \arg\max_T \; p(t \mid T)\, p(T)$$

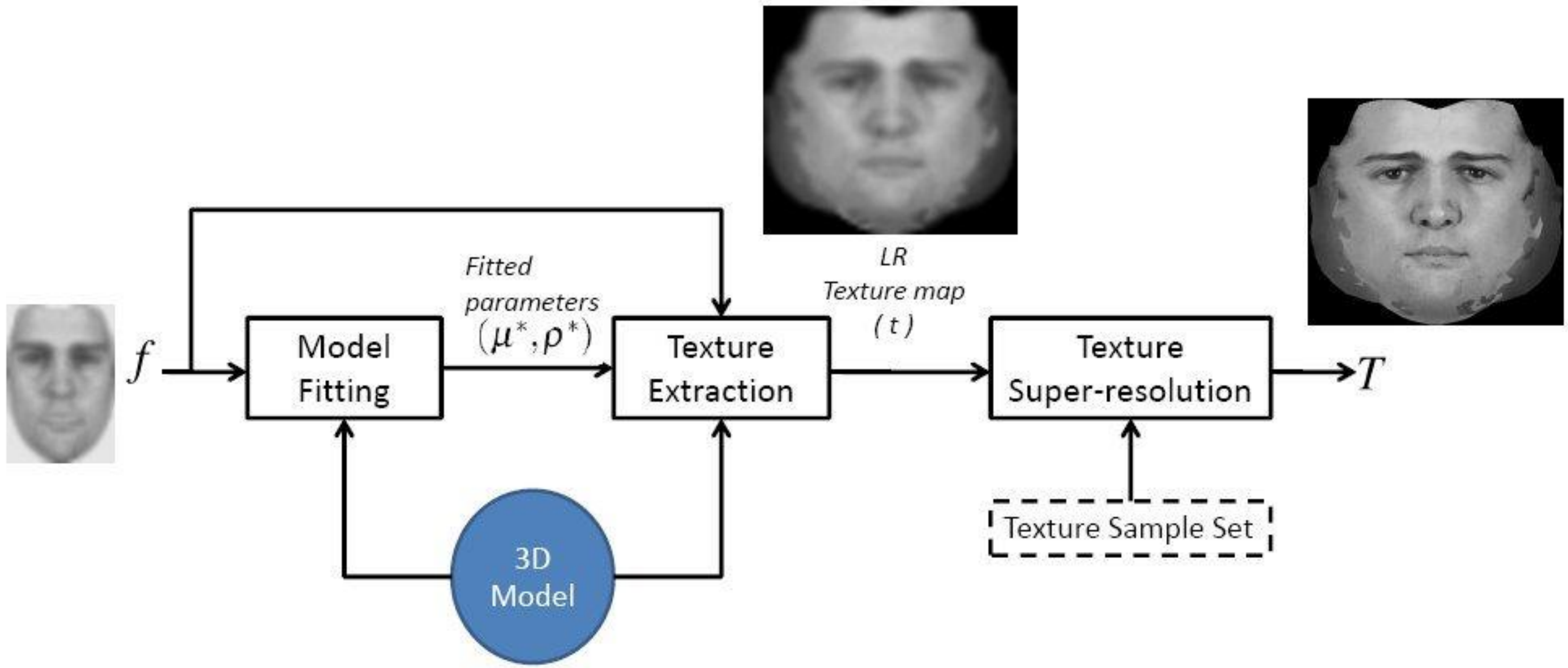
3D-Assisted SR:



3D-Assisted SR:



3D-Assisted SR:



- Likelihood: $-\log p(t \mid T) \sim \| AT - t \|$

- Likelihood: $$-\log p(t \mid T) \sim \| AT - t \|$$

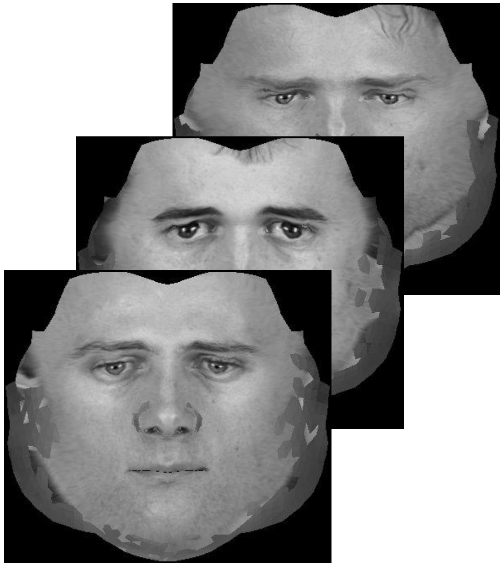
- Prior:

  Gradient Prediction:

• Likelihood: $$-\log p(t \mid T) \sim \| AT - t \|$$

• Prior:

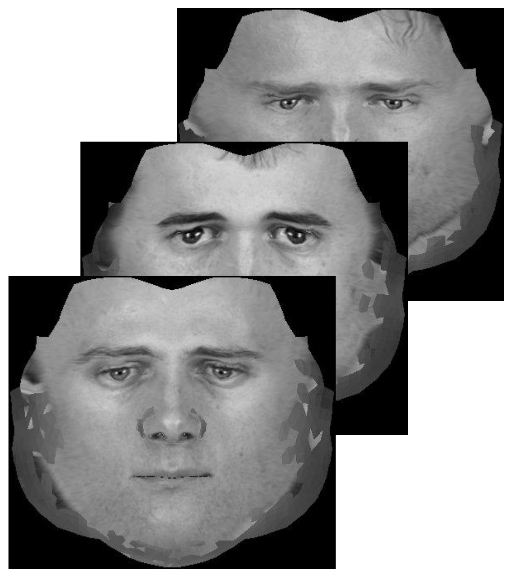
Gradient Prediction:



• Likelihood: $$-\log p(t \mid T) \sim \| AT - t \|$$

• Prior:

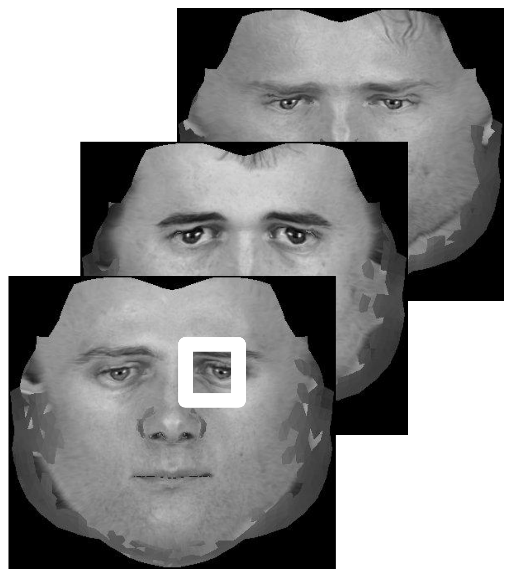
Gradient Prediction:



• Likelihood: $$-\log p(t \mid T) \sim \| AT - t \|$$

• Prior:

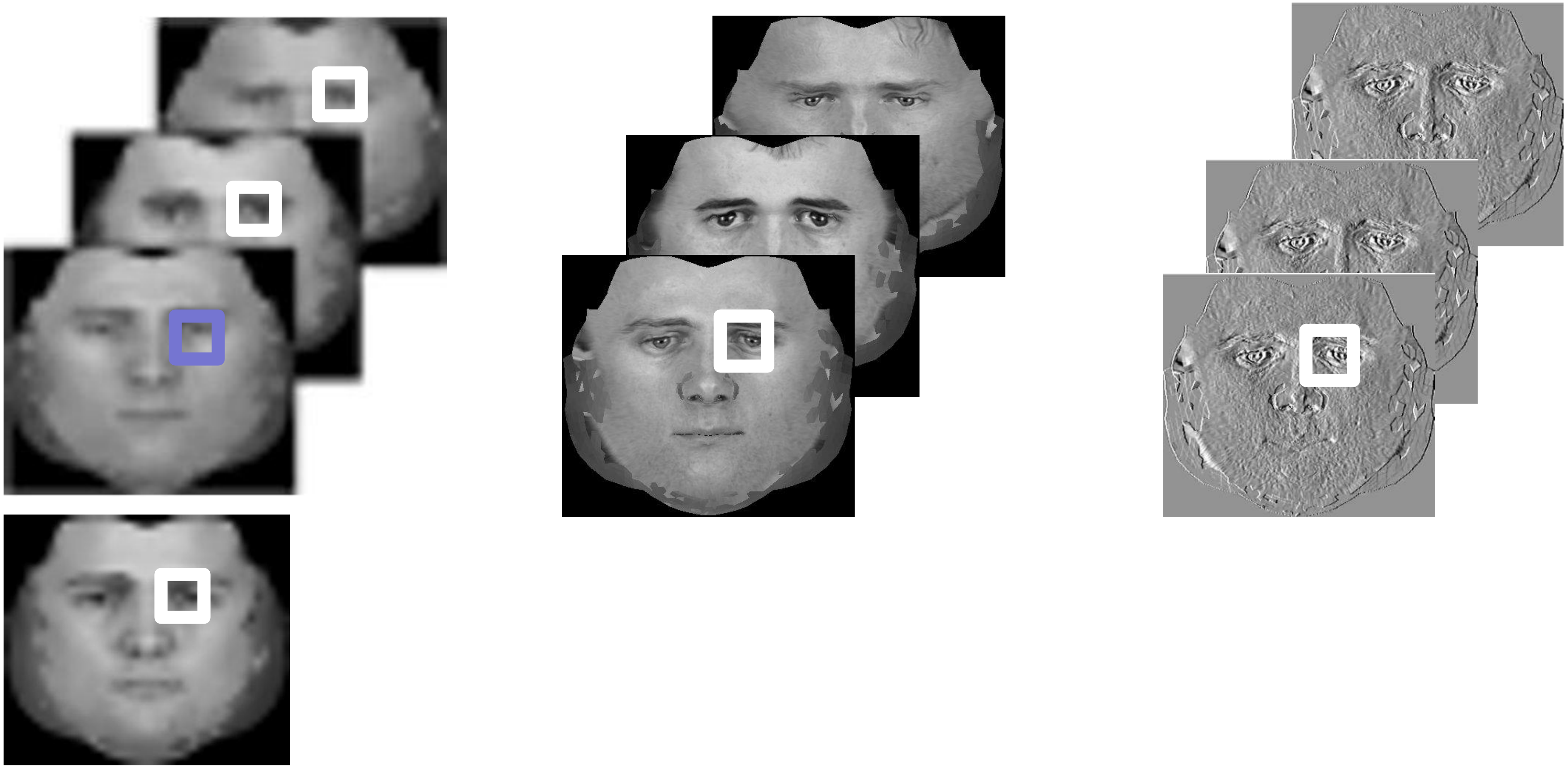
Gradient Prediction:



• Likelihood: $$-\log\, p(t \mid T) \sim \| AT - t \|$$

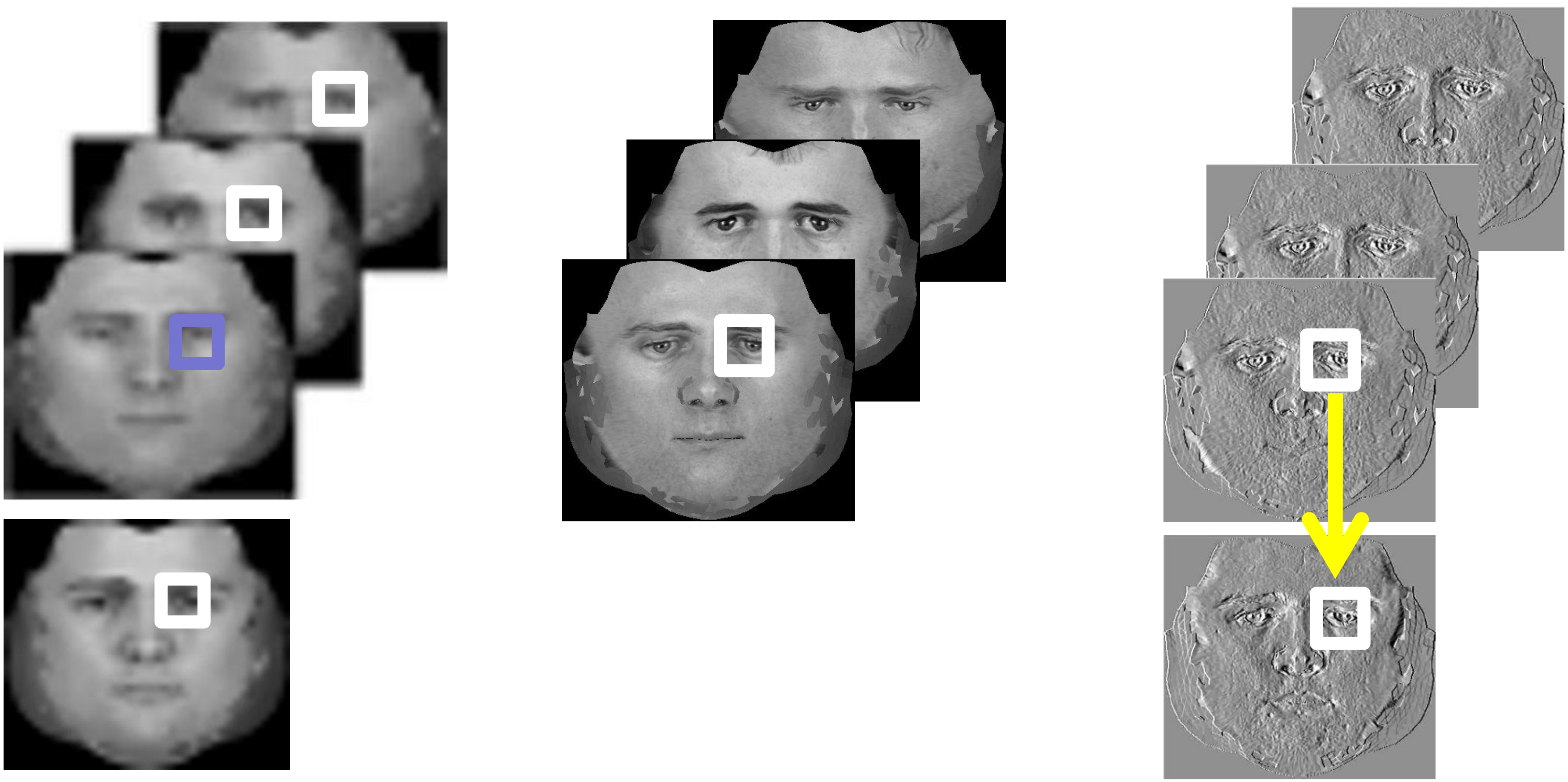
• Prior:

Gradient Prediction:



- Likelihood: $-\log p(t \mid T) \sim \| AT - t \|$

- Prior:
  Gradient Prediction:



- Likelihood: $$-\log p(t \mid T) \sim \| AT - t \|$$

- Prior:

  Gradient Prediction:

$$-\log p(T) \sim \sum_{m,n} [\hat{G}(m,n) - G(m.n)]^2$$

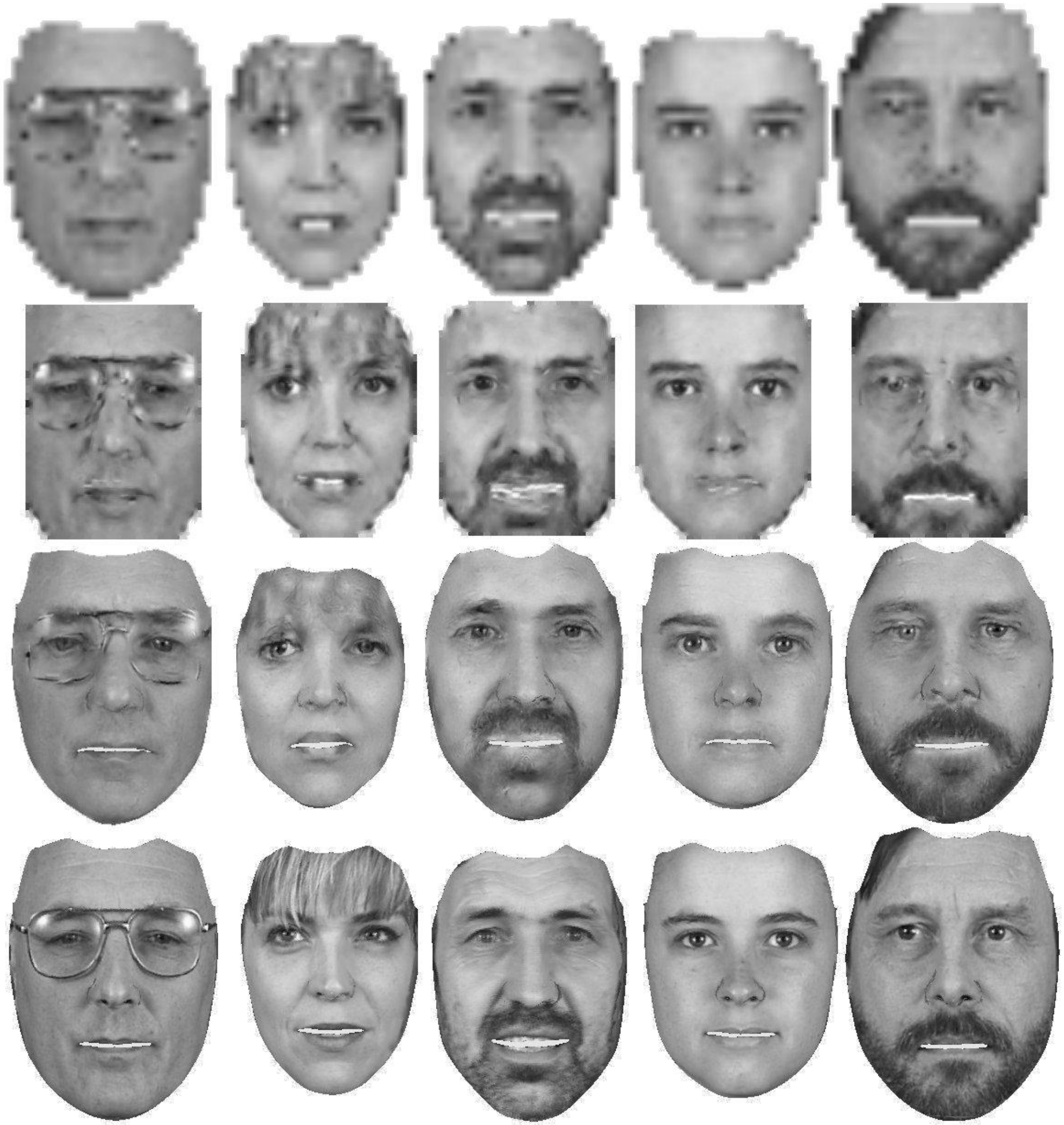
Results:

(a) (b) (C) (d) (e)

Results:

(a) (b) (C) (d) (e)

Results:

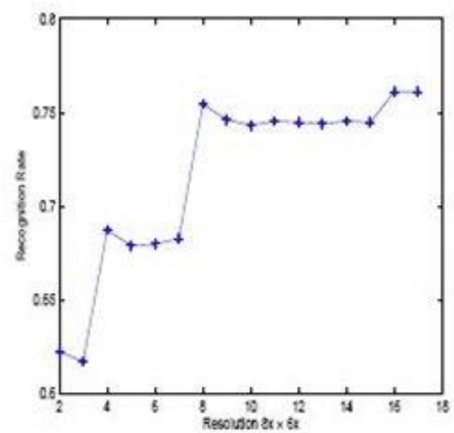


Results:

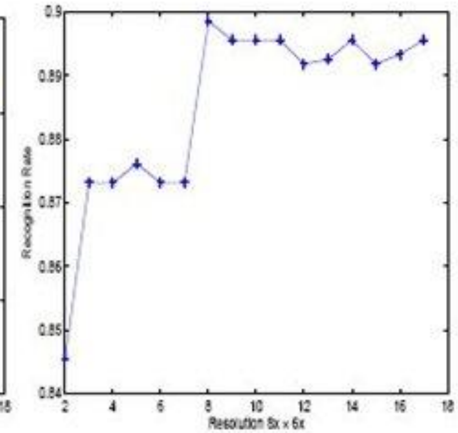
(a) (b) (c)

Results (Face Recognition):

Resolution of the input image can affect recognition performance



PCA

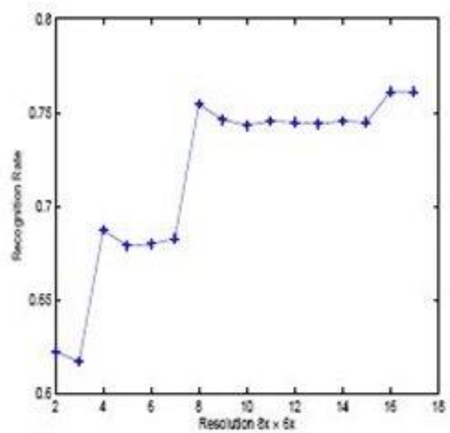


LDA

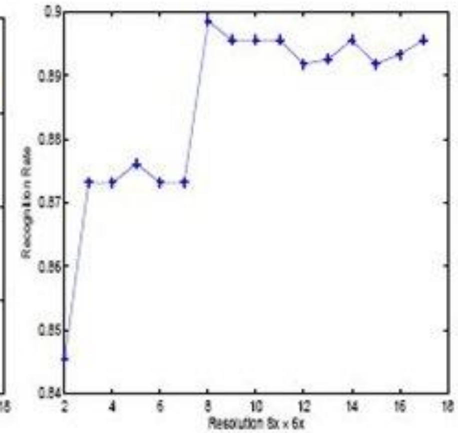
* *J. Wang, C. Zhang, H. Shum, " FACE IMAGE RESOLUTION VERSUS FACE RECOGNITION PERFORMANCE BASED ON TWO GLOBAL METHODS",* Proceedings of Asia Conference on Computer Vision (ACCV'04)

Results (Face Recognition):

Resolution of the input image can affect recognition performance



PCA



LDA

* *J. Wang, C. Zhang, H. Shum, " FACE IMAGE RESOLUTION VERSUS FACE RECOGNITION PERFORMANCE BASED ON TWO GLOBAL METHODS",* Proceedings of Asia Conference on Computer Vision (ACCV'04)

- XM2VTS
- LBP histograms + LDA
- Normalized Correlation
- 3 samples for training and 3 for test

FACE IDENTIFICATION RESULTS

Method	Identification Rate
HR	99.28
LR+Bilinear	78.57
LR+Baker_Kande	96.43
LR+Our Method	95

Conclusions:

- Our framework can deal with pose-independent face super-resolution.

- The results obtained are visually comparable to Face Hallucination in the image domain.

- The proposed method can provide additional information for face recognition.

- Model fitting on low-resolution images is not ideal and can degrade the results. However, its effect is not detrimental to the final result.

Thank You