Model-Based Face Recognition

Simon Baker
The Robotics Institute
Carnegie Mellon University

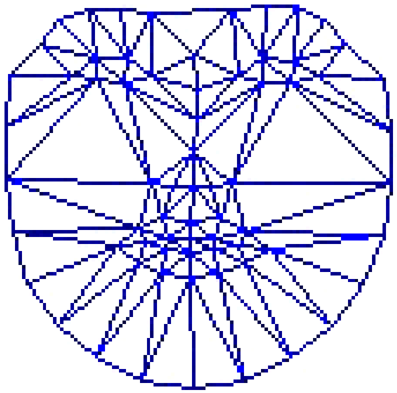
What is a Face Model?

Example Face Models

- 2D Active Appearance Models [Cootes and Taylor, 1998]
- 3D Morphable Models [Blanz and Vetter, 1999]

Appearance

Shape



Combined

AAM Above has 17 Shape and 42 Appearance Parameters
Movie created by varying first 4 (combined) parameters
Can Model: Identity, Pose, Illumination, Expresssion, …

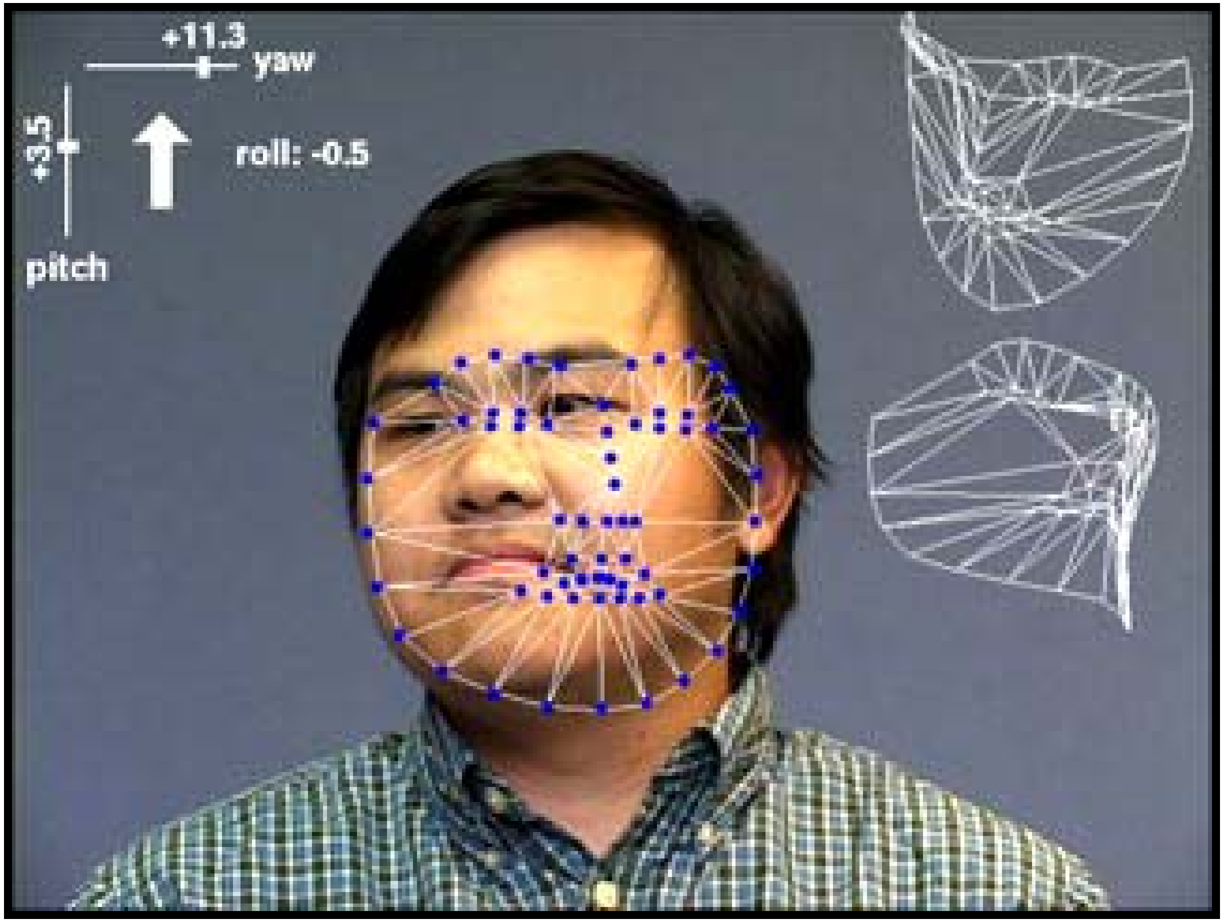
Fitting a Face Model

Fitting = Inverting Face Model "Function"

Fitting = Non-Linear Optimization

Fitting Example (3D)

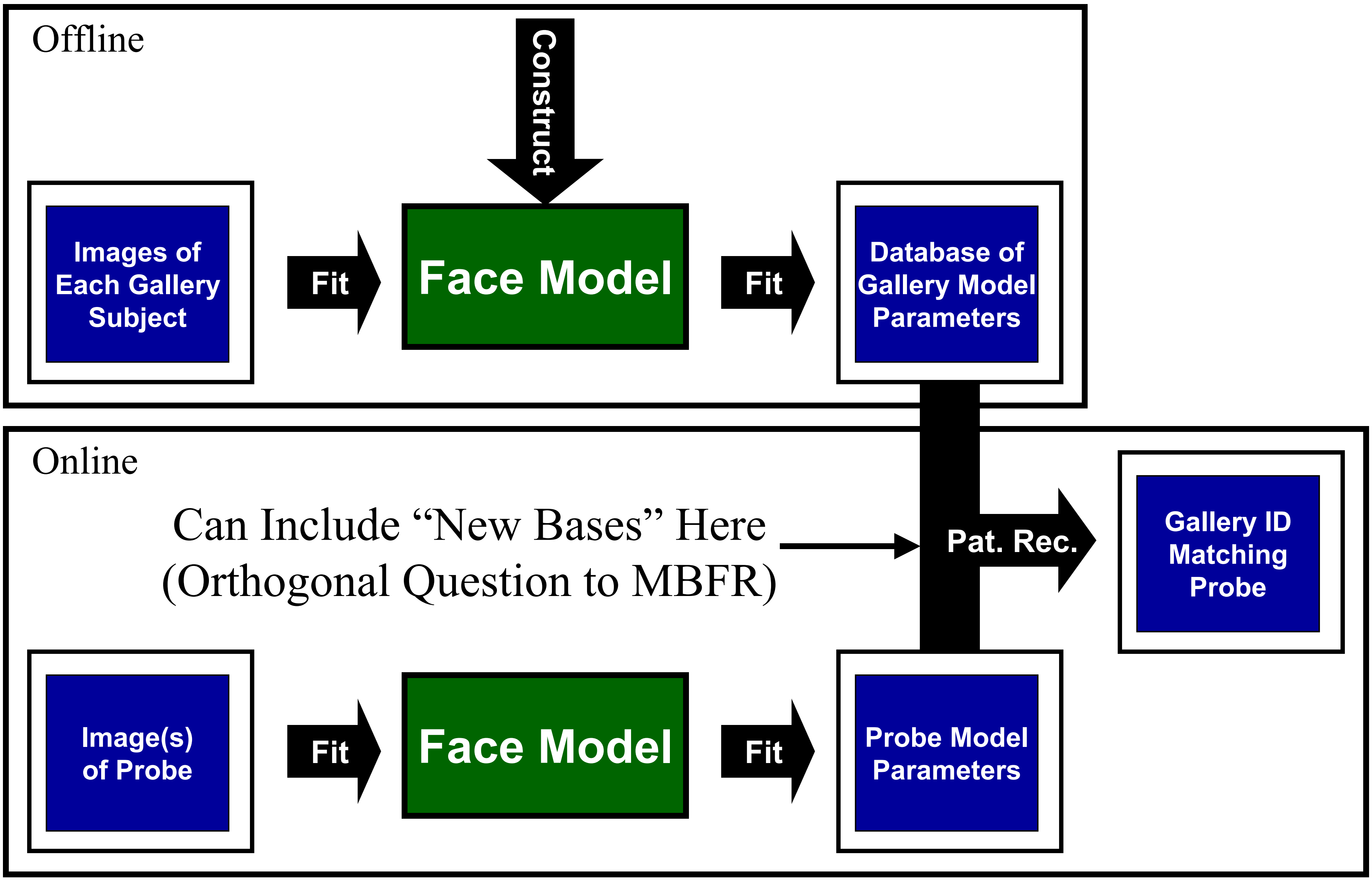
3D Pose



3D Shape

Computation Time ~ 4 milliseconds (~ 250Hz)
[Xiao, Baker, Matthews, and Kanade, 2004]

Model-Based Face Recognition Paradigm

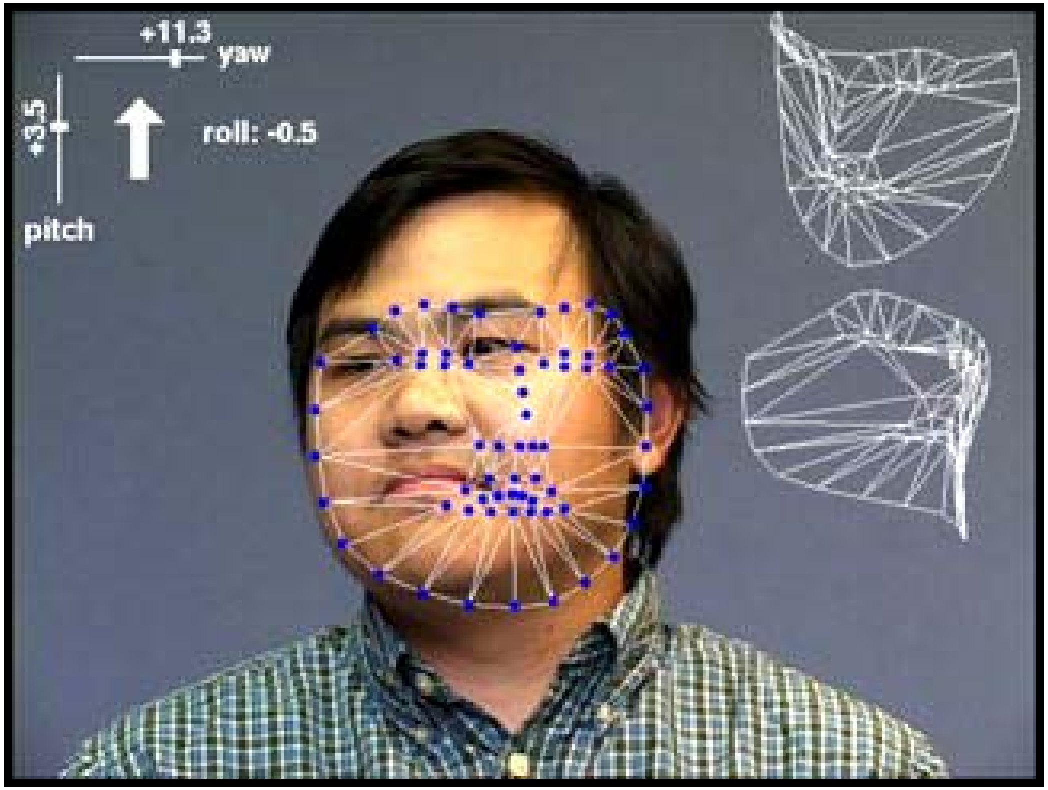


Why Model-Based FR?

- Jonathon Philips: “3DMMs Helped a Lot”

Why Model-Based FR? - Registration

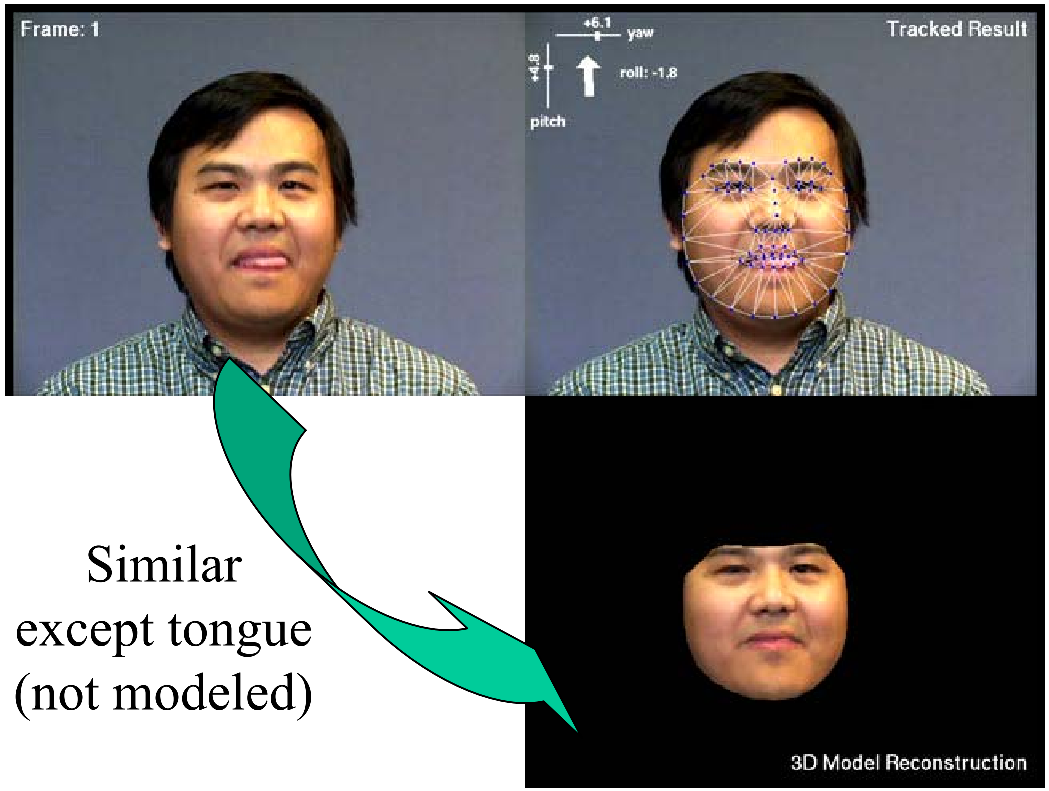
- Takeo Kanade: “Registration Most Important Task for FR”
- Model Fitting = Registration (of Eyes, Nose, Mouth, etc)



Why MBFR? – Efficient Non-Linear Coding

- Good Reconstruction => Parameters Contain Identity Info.

Input Image



Model Fit

Reconstructed From Parameters

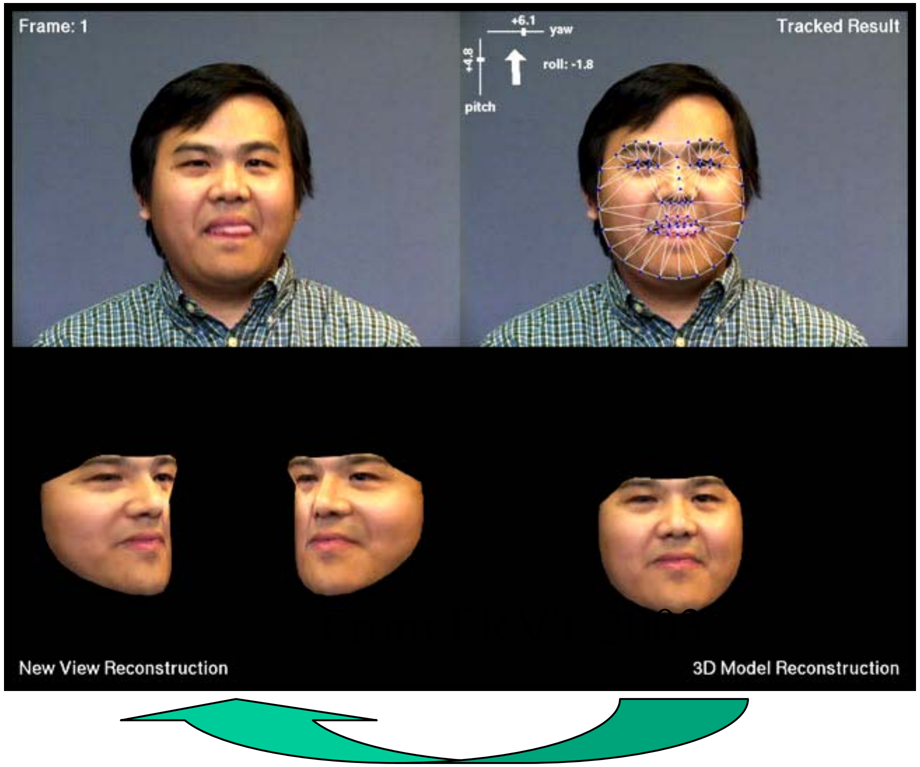
- Note: Reconstruction far better than Eigenfaces, etc
  - **(Can Do LDA, Different Bases, etc, on Parameters For Discrimination)**
- Also: Less Parameters => “Pattern Recognition” Easier

Why MBFR? – FRAP/Pose Normalization

- Face Recognition Across Pose/Pose Normalization:
  - By Far Best Results to Date are Model Based: [Blanz et al., 2003]

Input Image

Model Fit



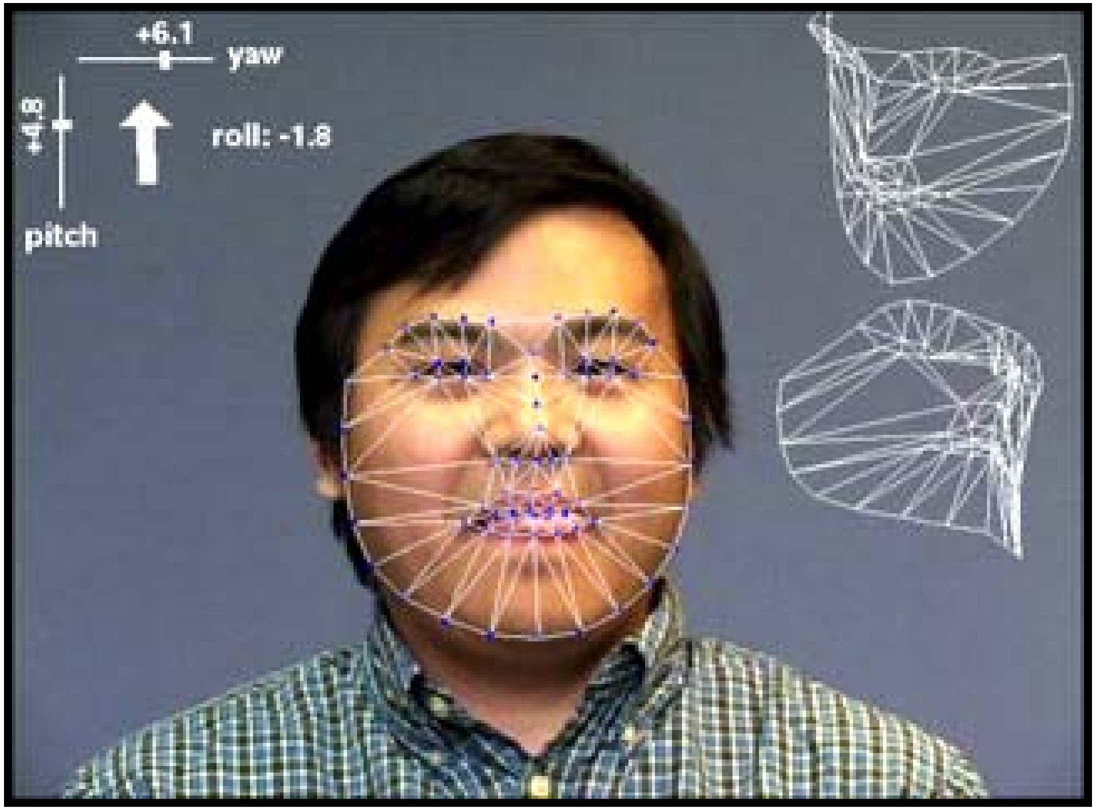
Reconstructed Novel Poses

Reconstructed Same Pose

- Better results by fitting model to >1 frames across pose
  - See [Gross, Matthews, and Baker, 2004]

Why Model-Based FR? - Video

- Model Fitting to Sequential Frames = Non-Rigid Tracking
  - Required for FR from expression, deception-detection, etc

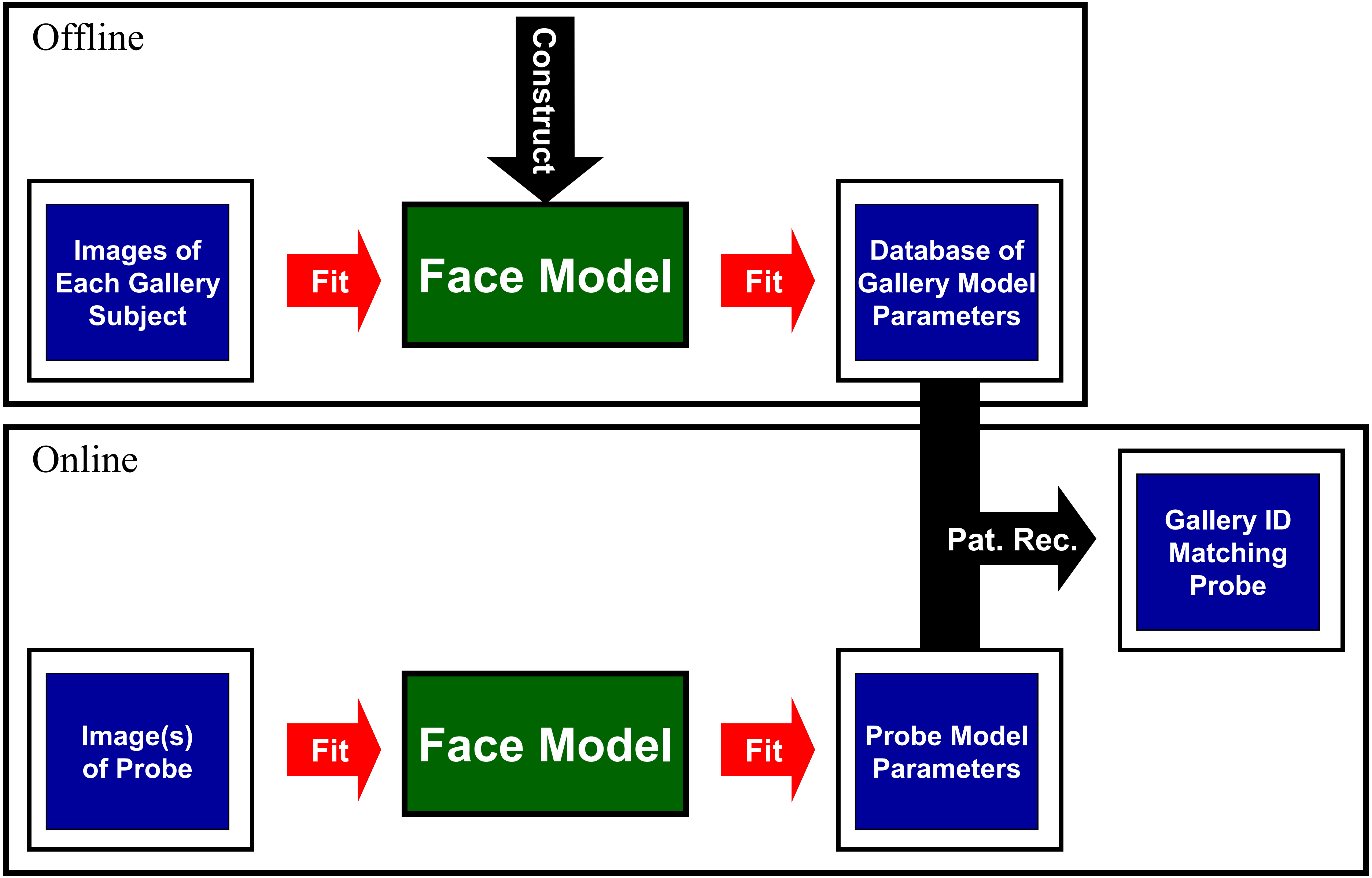


- May get better FR results by fitting model to entire video

Why is MBFR Not Used in Real Systems?

- Model Fitting: Existing Algorithms:
  - Are Too slow
  - Don’t work very well (fast algorithms make strong assumptions)
  - Are not very easy to implement
- Model Construction:
  - Very time consuming (substantial user input required)
  - 3D models require range data (hard to obtain)
- Most MBFR Research in Europe (ECCV/BMVC/…)?

CMU-RI Model Fitting Research



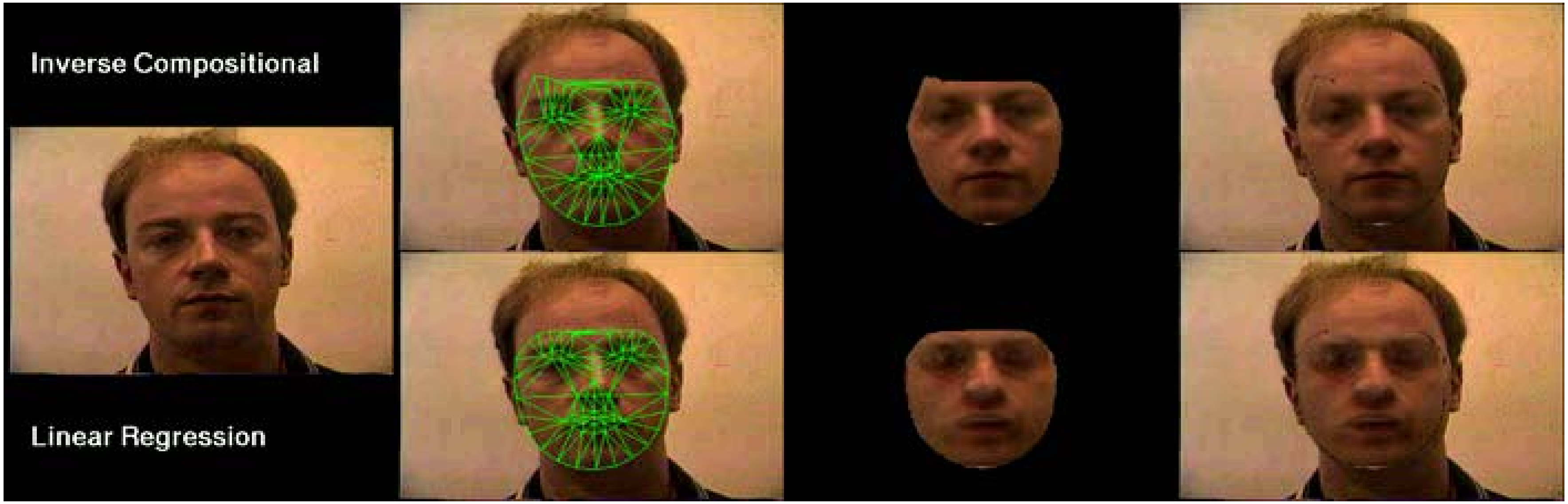
Previous Work: 2D AAM Fitting Algorithm

- Analytically Derived Efficient Gradient Descent Algorithm
  - 230 Frames Per Second

- Papers:
  - Original Paper [Baker and Matthews, CVPR, 2001]
  - Inverse Compositional Algorithm [Baker and Matthews, IJCV, 2004]
  - Application to AAMs [Matthews and Baker, IJCV, 2004]

Comparison of 2D AAM Algorithms

- Top: Our Algorithm: [Matthews and Baker, IJCV, 2004]
  - Analytically derived gradient descent, 230 frames per second



- Bottom: Original AAM Algorithm [Cootes and Taylor, 1998]
  - Numerical Algorithm, A few (5-10) frames per second
- Which would result in better Face Recognition?

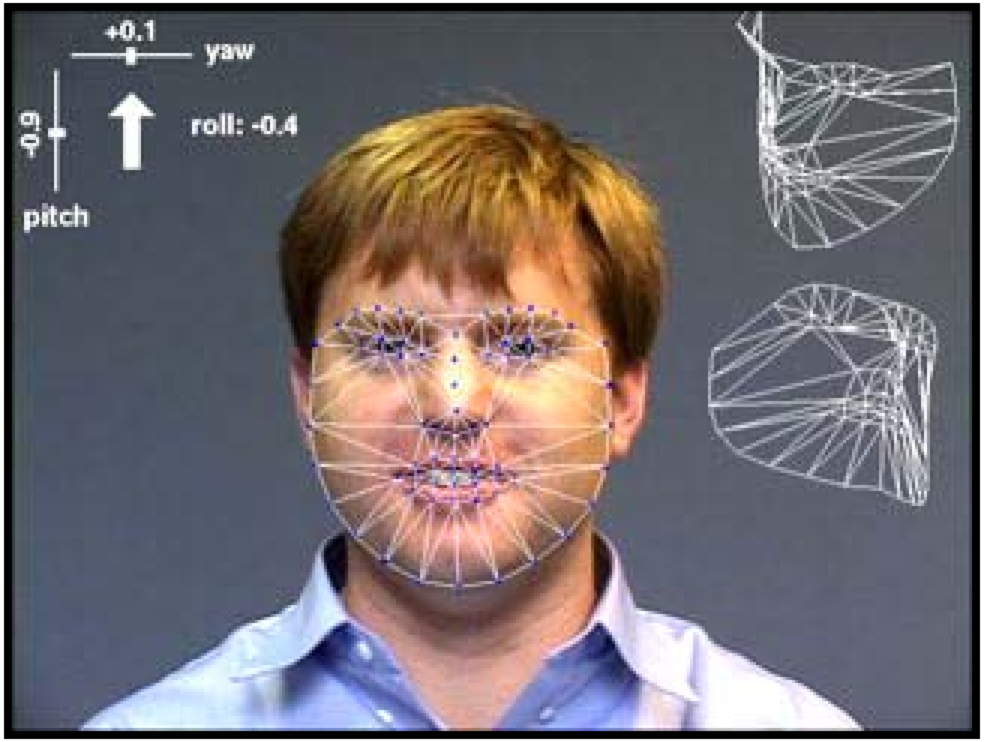
Previous 3D Fitting Algorithms

- Original 3D Morphable Models Alg. [Blanz & Vetter, 1999]
  - Stochastic Gradient Descent Algorithm
  - Requires hand-initialization of 7 points
  - Approximately 2 minutes per frame
- "Efficient" Algorithm [Romdhami and Vetter, 2003]
  - Based on "Inverse Compositional" Alg. [Baker and Matthews, 2004]
  - Still approximately 30 seconds per frame
- Too Slow for Incorporation into Systems Today…

CMU-RI 3D AAM Fitting Algorithm

- Extension of (230Hz) 2D Algorithm
  - 286 Frames Per Second (faster b/c 3D more constrained problem)

Recovered 3D Pose



Recovered 3D Shape

- Papers (Under Review):
  - 3D Algorithm: [Xiao, Baker, Matthews, and Kanade, CVPR, 2004]
  - Allowing Occlusion: [Gross, Matthews, and Baker, CVPR, 2004]

How Does It Work? – 2D Fitting Goal

- 2D Fitting Goal:

Appearance Parameters   Shape Variation+Parameters

$$\arg\min_{\mathbf{p},\lambda} \sum_{\mathbf{x}} \left[\boxed{T}(\mathbf{x}) + \sum_{i} \boxed{\lambda_i}\,\boxed{A_i}(\mathbf{x}) - \boxed{I}(\boxed{\mathbf{W}}(\mathbf{x};\boxed{\mathbf{p}})) \right]^2$$

“Mean” Face   Appearance Variation   Input Image

- Previous Work:
  - Efficient (230Hz) 2D Inverse Compositional Algorithm
  - See [Matthews and Baker, 2004]

How Does It Work? – 2D Fitting Goal

- Fitting Goal:

Appearance Parameters   Shape Variation+Parameters

$$\arg\min_{\mathbf{p},\lambda} \sum_{\mathbf{x}} \left[\boxed{T}(\mathbf{x}) + \sum_i \boxed{\lambda_i}\,\boxed{A_i}(\mathbf{x}) - \boxed{I}(\boxed{\mathbf{W}}(\mathbf{x};\boxed{\mathbf{p}}))\right]^2$$

“Mean” Face   Appearance Variation   Input Image

- Ignoring Appearance (for now):

$$\arg\min_{\mathbf{p}} \sum_{\mathbf{x}} \left[T(\mathbf{x}) - I(\mathbf{W}(\mathbf{x};\mathbf{p}))\right]^2$$

Standard Gauss-Newton Gradient Descent

- Assume **current p** is fixed, solve for **additive** Δ**p**:

$$\arg\min_{\Delta\mathbf{p}} \sum_{\mathbf{x}} \left[T(\mathbf{x}) - I(\mathbf{W}(\mathbf{x};\mathbf{p}+\Delta\mathbf{p})) \right]^2 \qquad \mathbf{p} \leftarrow \mathbf{p} + \Delta\mathbf{p}$$

- Gauss-Newton First-Order Least-Squares Solution:

$$\Delta\mathbf{p} = \boxed{H^{-1}} \sum_{\mathbf{x}} \boxed{\left[\nabla I \left.\frac{\partial \mathbf{W}}{\partial \mathbf{p}}\right|_{\mathbf{p}} \right]^{\mathrm{T}}} \left[T(\mathbf{x}) - I(\mathbf{W}(\mathbf{x};\mathbf{p})) \right]$$

$$\boxed{H = \sum_{\mathbf{x}} \left[\nabla I \left.\frac{\partial \mathbf{W}}{\partial \mathbf{p}}\right|_{\mathbf{p}} \right]^{\mathrm{T}} \left[\nabla I \left.\frac{\partial \mathbf{W}}{\partial \mathbf{p}}\right|_{\mathbf{p}} \right]}$$

Depends on **p** & I
Must Update
Slow to Compute

Equivalent Gradient Descent Formulations

- Usual Additive Formulation:

$$\arg\min_{\Delta\mathbf{p}} \sum_{\mathbf{x}} \left[T(\mathbf{x}) - I(\mathbf{W}(\mathbf{x};\mathbf{p}+\Delta\mathbf{p})) \right]^2$$

$$\mathbf{p} \leftarrow \mathbf{p} + \Delta\mathbf{p}$$

- Compositional Formulation:

$$\arg\min_{\Delta\mathbf{p}} \sum_{\mathbf{x}} \left[T(\mathbf{x}) - I(\mathbf{W}(\mathbf{W}(\mathbf{x};\Delta\mathbf{p});\mathbf{p})) \right]^2$$

$$\mathbf{W}(\mathbf{x};\mathbf{p}) \leftarrow \mathbf{W}(\mathbf{x};\mathbf{p}) \circ \mathbf{W}(\mathbf{x};\Delta\mathbf{p})$$

More Equivalent Formulations

- Forwards Compositional Formulation:

$$\arg\min_{\Delta\mathbf{p}} \sum_{\mathbf{x}} \left[T(\mathbf{x}) - I(\mathbf{W}(\mathbf{W}(\mathbf{x};\Delta\mathbf{p});\mathbf{p})) \right]^2$$

$$\mathbf{W}(\mathbf{x};\mathbf{p}) \leftarrow \mathbf{W}(\mathbf{x};\mathbf{p}) \circ \mathbf{W}(\mathbf{x};\Delta\mathbf{p})$$

- Inverse Compositional Formulation:

$$\arg\min_{\Delta\mathbf{p}} \sum_{\mathbf{x}} \left[T(\mathbf{W}(\mathbf{x};\Delta\mathbf{p})) - I(\mathbf{W}(\mathbf{x};\mathbf{p})) \right]^2$$

$$\mathbf{W}(\mathbf{x};\mathbf{p}) \leftarrow \mathbf{W}(\mathbf{x};\mathbf{p}) \circ \mathbf{W}(\mathbf{x};\Delta\mathbf{p})^{-1}$$

Inverse Compositional Derivation

Inverse Compositional Formulation:

$$\arg\min_{\Delta\mathbf{p}} \sum_{\mathbf{x}} \left[T(\mathbf{W}(\mathbf{x};\mathbf{0}+\Delta\mathbf{p})) - I(\mathbf{W}(\mathbf{x};\mathbf{p})) \right]^2$$

$$\mathbf{W}(\mathbf{x};\mathbf{p}) \leftarrow \mathbf{W}(\mathbf{x};\mathbf{p}) \circ \mathbf{W}(\mathbf{x};\Delta\mathbf{p})^{-1}$$

- First-Order Least-Squares Solution:

$$\Delta\mathbf{p} = -\boxed{H^{-1}\sum_{\mathbf{x}}} \boxed{\left[\nabla T \frac{\partial \mathbf{W}}{\partial \mathbf{p}}\bigg|_{\mathbf{0}} \right]^{\mathrm{T}}} \left[T(\mathbf{x}) - I(\mathbf{W}(\mathbf{x};\mathbf{p})) \right]$$

$$\boxed{H = \sum_{\mathbf{x}} \left[\nabla T \frac{\partial \mathbf{W}}{\partial \mathbf{p}}\bigg|_{\mathbf{0}} \right]^{\mathrm{T}} \left[\nabla T \frac{\partial \mathbf{W}}{\partial \mathbf{p}}\bigg|_{\mathbf{0}} \right]}$$

Constant

Can Pre-Compute

Appearance – Use Orthogonality

$$\sum_{\mathbf{x}} \left[T(\mathbf{x}) + \sum_i \lambda_i A_i(\mathbf{x}) - I(\mathbf{W}(\mathbf{x};\mathbf{p})) \right]^2 =$$

$$\boxed{\left\| T(\mathbf{x}) + \sum_i \lambda_i A_i(\mathbf{x}) - I(\mathbf{W}(\mathbf{x};\mathbf{p})) \right\|_{\mathrm{span}(A_i)}} +$$

argmin = 0
Always !!

$$\boxed{\left\| T(\mathbf{x}) + \sum_i \lambda_i A_i(\mathbf{x}) - I(\mathbf{W}(\mathbf{x};\mathbf{p})) \right\|_{\mathrm{span}(A_i)^\perp}}$$

$$\boxed{= \left\| T(\mathbf{x}) - I(\mathbf{W}(\mathbf{x};\mathbf{p})) \right\|_{\mathrm{span}(A_i)^\perp}}$$

How Does It Work? – Appearance

$$\arg\min_{\mathbf{p},\lambda} \sum_{\mathbf{x}} \left[T(\mathbf{x}) + \sum_i \lambda_i A_i(\mathbf{x}) - I(\mathbf{W}(\mathbf{x};\mathbf{p})) \right]^2$$

Step 1: $\arg\min_{\mathbf{p}} \left\| T(\mathbf{x}) - I(\mathbf{W}(\mathbf{x};\mathbf{p})) \right\|_{\mathrm{span}(A_i)^{\perp}}$

(Step 1: No Appearance)

Step 2:

$$\arg\min_{\lambda} \left\| T(\mathbf{x}) + \sum_i \lambda_i A_i(\mathbf{x}) - I(\mathbf{W}(\mathbf{x};\mathbf{p})) \right\|_{\mathrm{span}(A_i)}$$

(Step 2: Linear Closed form Solution, Essentially EigenFaces)

Extension of 2D Algorithm to 2D+3D

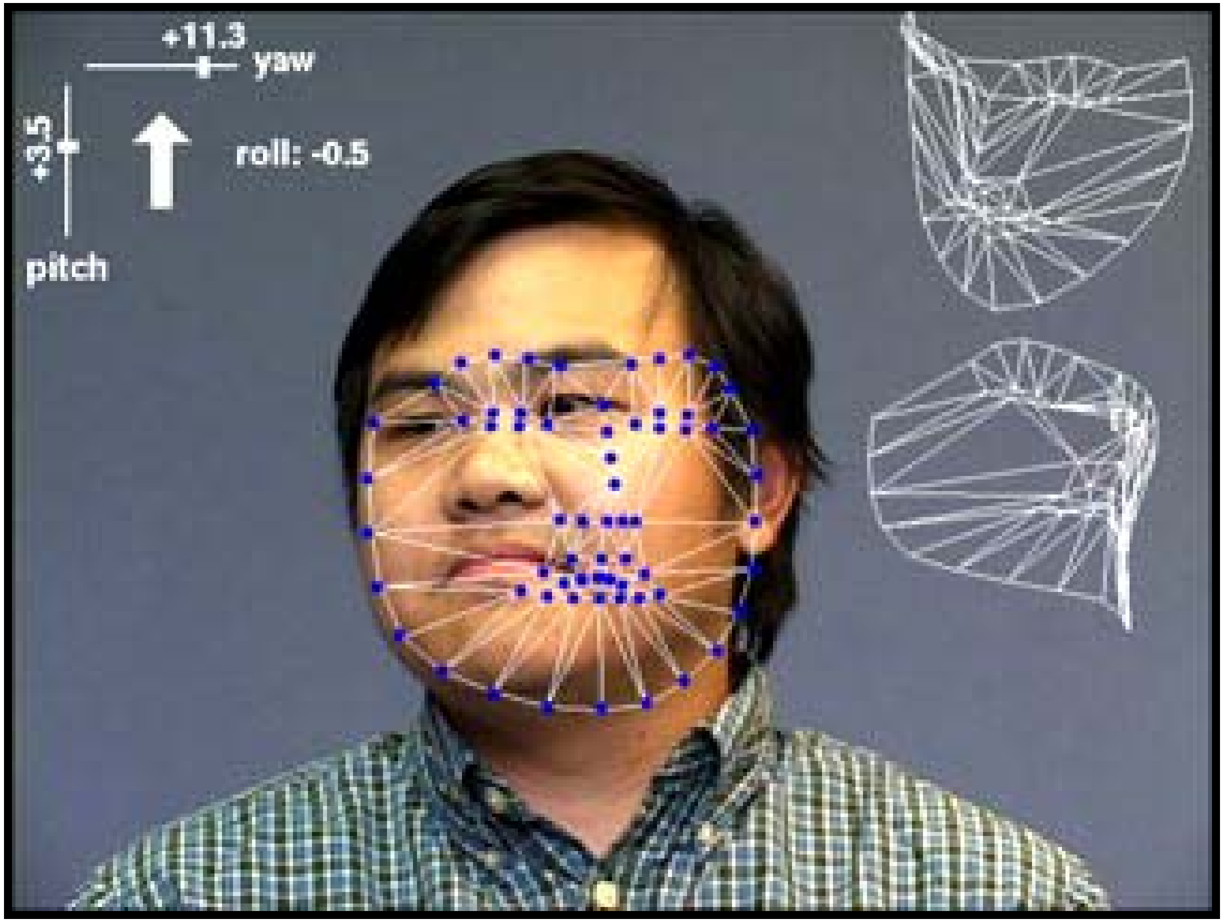
$$\arg\min_{\mathbf{p},\lambda} \sum_{\mathbf{x}} \left[T(\mathbf{x}) + \sum_i \lambda_i A_i(\mathbf{x}) - I(\mathbf{W}(\mathbf{x};\mathbf{p})) \right]^2$$

$$\arg\min_{\mathbf{p},\lambda,\mathbf{q},\mathbf{r}} \sum_{\mathbf{x}} \left[T(\mathbf{x}) + \sum_i \lambda_i A_i(\mathbf{x}) - I(\mathbf{W}(\mathbf{x};\mathbf{p})) \right]^2 + \boxed{F^2(\mathbf{p},\mathbf{q},\mathbf{r})}$$

Constraint: 2D shape **p** is projection of 3D shape **q** at pose **r**
Enforcing Constraint: Solves for 3D shape **q** and 3D pose **r**
F(**p, q, r**) does not depend on image so computing it is quick

Recap: Fitting Example (2D+3D AAM)

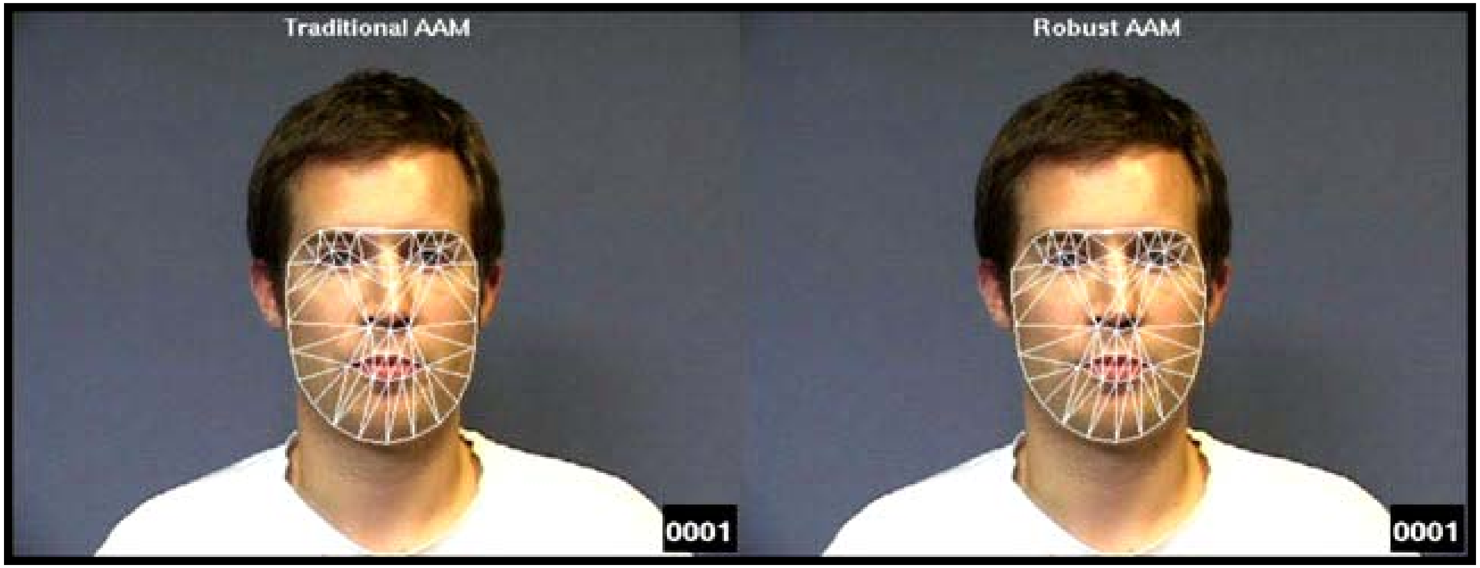
3D Pose **r**



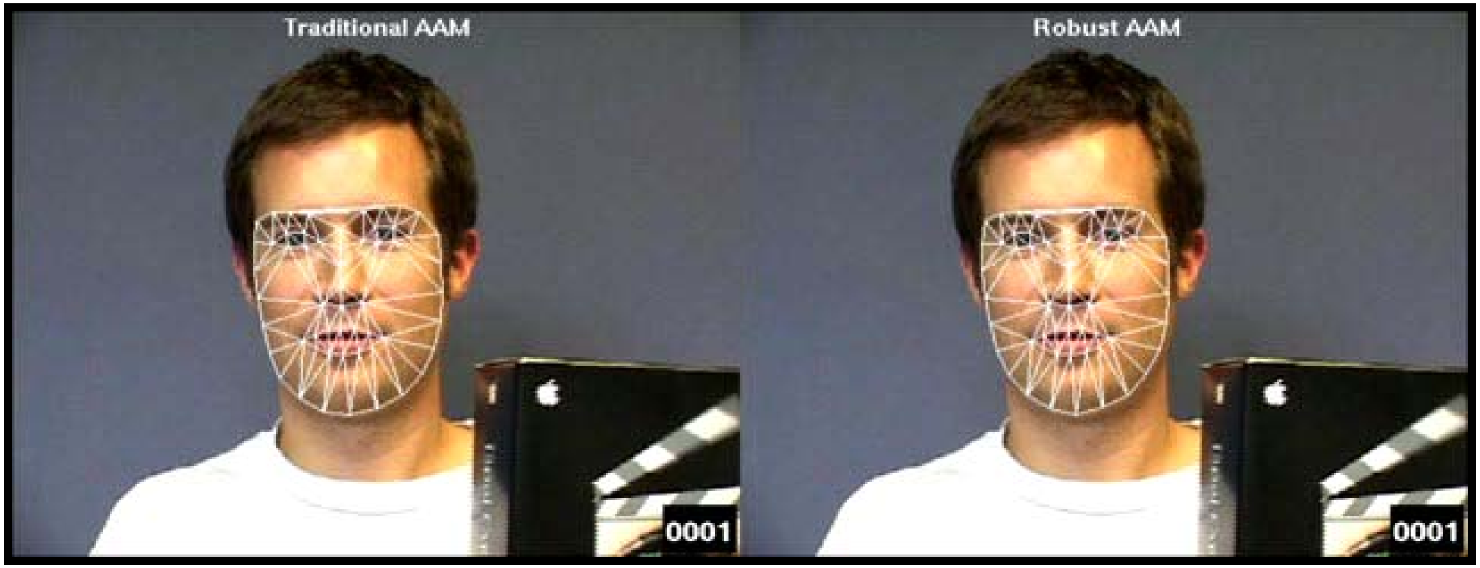
3D Shape **q**

Blue Dots: 2D Shape **p**
White Lines: 3D Shape **q**

Robustness to Pose Variation/Occlusion



Without Occlusion Modeling | With Occlusion Modeling



Occlusion Modeling

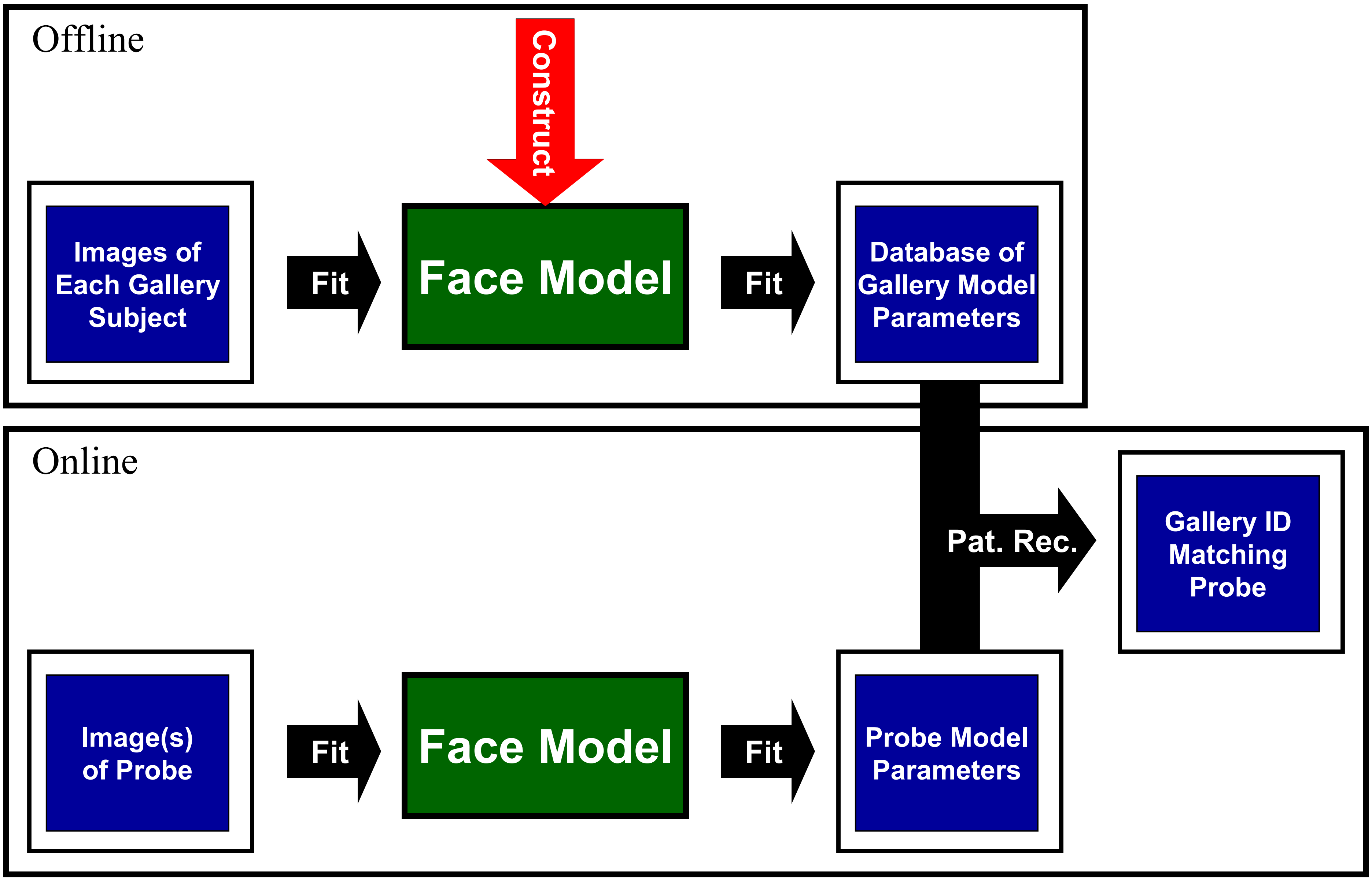
$$\arg\min_{\mathbf{p},\lambda,\mathbf{q},\mathbf{r}} \sum_{\mathbf{x}} \left[T(\mathbf{x}) + \sum_i \lambda_i A_i(\mathbf{x}) - I(\mathbf{W}(\mathbf{x};\mathbf{p})) \right]^2 + F^2(\mathbf{p},\mathbf{q},\mathbf{r})$$

$$\Downarrow$$

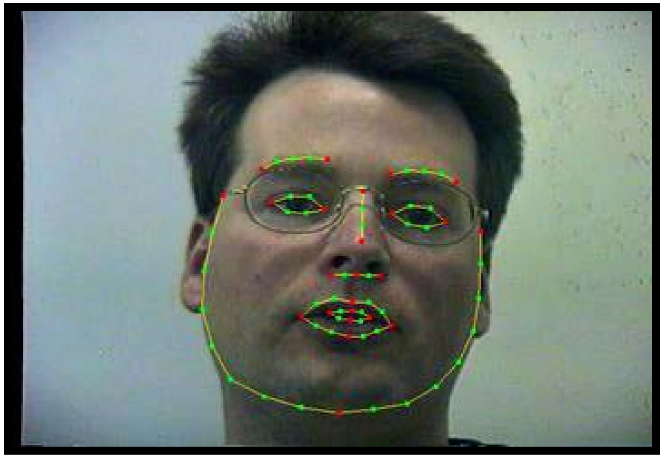
$$\arg\min_{\mathbf{p},\lambda,\mathbf{q},\mathbf{r}} \sum_{\mathbf{x}} \boxed{\rho} \left[T(\mathbf{x}) + \sum_i \lambda_i A_i(\mathbf{x}) - I(\mathbf{W}(\mathbf{x};\mathbf{p})) \right] + F^2(\mathbf{p},\mathbf{q},\mathbf{r})$$

Simply add a Robust Norm, right?
Ok, but keeping the algorithm efficient requires some work…
For the details, see [Gross, Matthews, and Baker, 2004] or the TRs:
"Lucas-Kanade 20 Years On: A Unifying Framework: Parts 2&3"

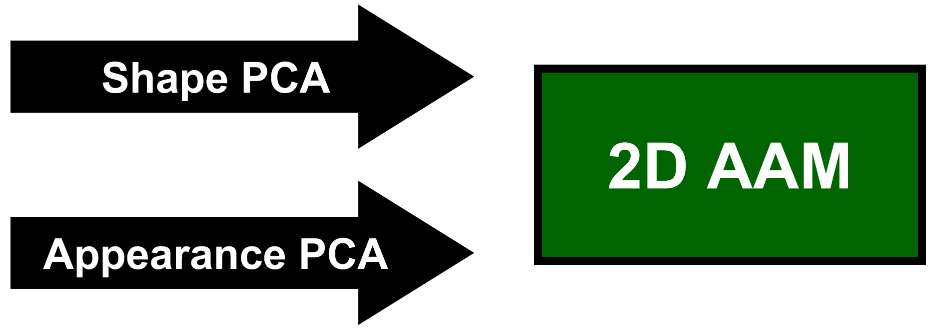
Model Construction Research



Model Construction (Offline)



Hand-Labeled Training Data



- (Selected) CMU-RI Work On Model Construction:
  - Avoiding Hand Labeling: [Baker, Matthews, and Schneider, 2003]
  - 3D AAMs from 2D Images: [Xiao, Baker, Matthews, & Kanade, 2004]

Progress Towards Using MBFR in Industry

- Model Fitting:
  - Real-time algorithms for 2D and 3D models (>200Hz)
  - Analytically derived algorithms
  - Happy to work with/provide implementation to industry
    - (Although not making code freely available like PIE DB/LK20 code)

- Model Construction:
  - Some progress on automatic construction (not a total solution)
  - (Non-rigid SFM) algorithm for 3D models from 2D images

Ongoing Work

- Model Fitting and Construction:
  - Better Fitting Algorithms for Generic (Multi-Person) AAMs
  - Local Minima: Combine 286Hz Algorithm w/ Combinatorial Search
  - Algorithms for Optimizing the Compactness of the Model
- Face Recognition Across Pose (FRAP):
  - Develop “Multi-View AAMs” and a Fitting Algorithm For Them
  - Complete Collection of Multi-PI DB and Evaluate FRAP Algorithm
- Face Recognition From Video (FRFV):
  - Develop Algorithm to Fit an AAM Simultaneously to an Entire Video
  - Collect Face Video DB and Evaluate FRFV Algorithm

For More Details, Either Speak to Me Later Today/Tomorrow, or See My Webpage:

http://www.cs.cmu.edu/~simonb

Selected Papers Published/Submitted

- “The Template Update Problem,” Iain Matthews, Takahiro Ishikawa, and Simon Baker, BMVC, 2003.
- “Automatic Construction of Active Appearance Models as an Image Coding Problem,” Simon Baker, Iain Matthews, and Jeff Schneider, Submitted to PAMI, 2003.
- “Lucas Kanade 20 Years On: A Unifying Framework,” Simon Baker and Iain Matthews, Accepted To Appear in IJCV, 2004.
- “Active Appearance Models Revisited,” Iain Matthews and Simon Baker, Accepted To Appear in IJCV, 2004.
- “Appearance-Based Face Recognition and Light-Fields,” Ralph Gross, Iain Matthews, and Simon Baker, Accepted To Appear in PAMI, 2004.
- “Real-Time Combined 2D+3D Active Appearance Models,” Jing Xiao, Simon Baker, Iain Matthews, and Takeo Kanade, Submitted to CVPR, 2004.
- “Constructing and Fitting Active Appearance Models With Occlusion,” Ralph Gross, Iain Matthews, and Simon Baker, Submitted to CVPR, 2004.