graphics and vision gravis University of Basel

2D Face Image Analysis

Probabilistic Morphable Models

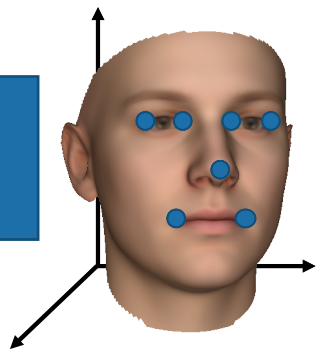
Summer School, June 2017

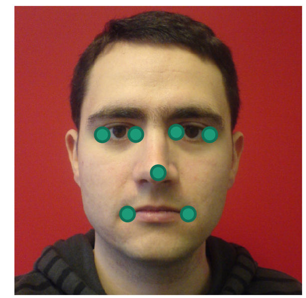
Sandro Schönborn

University of Basel

Contents

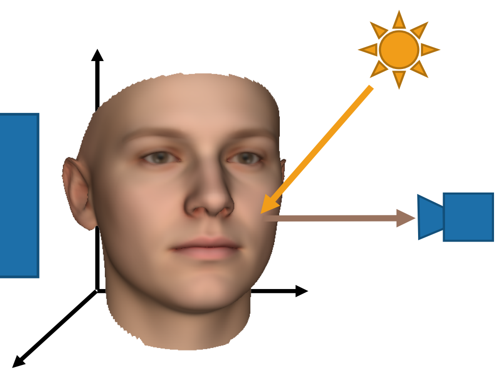
Landmarks Fitting

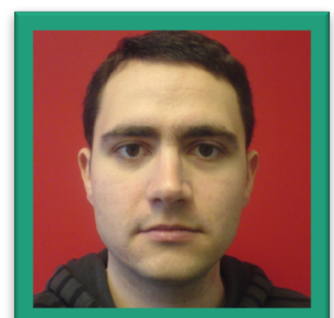




Observed Landmarks in 2D

Image Fitting



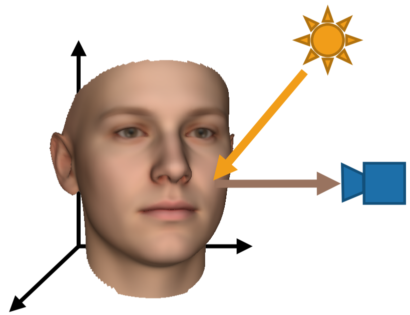


Observed Image

2D Face Image Analysis

Morphable Model adaptation to explain image

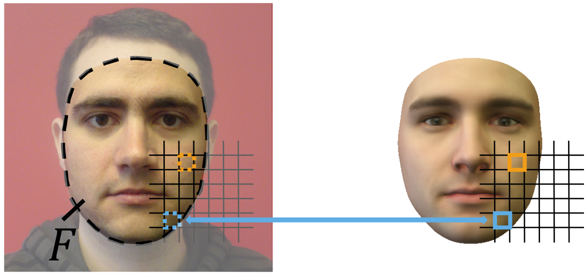
Bayesian Inference Setup



$$P(\theta|I) \propto \ell(\theta;I)P(\theta)$$

Image Likelihood

Image as observation



Face & Feature point detection

Fast bottom-up methods

3D Face Reconstruction

Fitting as Probabilistic Inference

- Probabilistic Inference Problem:

$$P(\theta|I) = \frac{P(I|\theta)P(\theta)}{N(I)} \qquad N(I) = \int P(I|\theta)P(\theta)\mathrm{d}\theta$$

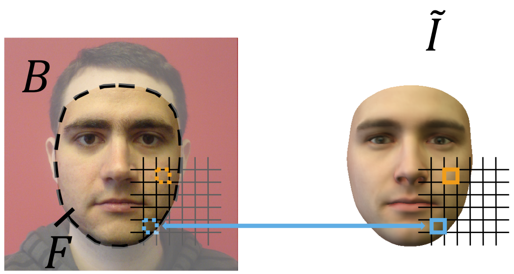
- *Prior*: $P(\theta)$
  Statistical face model

  Face shape & color (PPCA/GP models):

  $$s_\alpha = \mu + UD\alpha \qquad \alpha \sim N(0, I_d)$$

  Scene: illumination, pose, camera

- *Likelihood*: $P(I|\theta)$
  Image is *observation*



$$\ell(\theta; I) = \prod_{i \in F} \mathcal{N}\left(I_i \mid \tilde{I}_i(\theta), \sigma^2 I_3\right) \prod_{j \in B} b_{\mathrm{BG}}(I_i)$$

MH Inference of the 3DMM

- Target distribution is our "posterior":

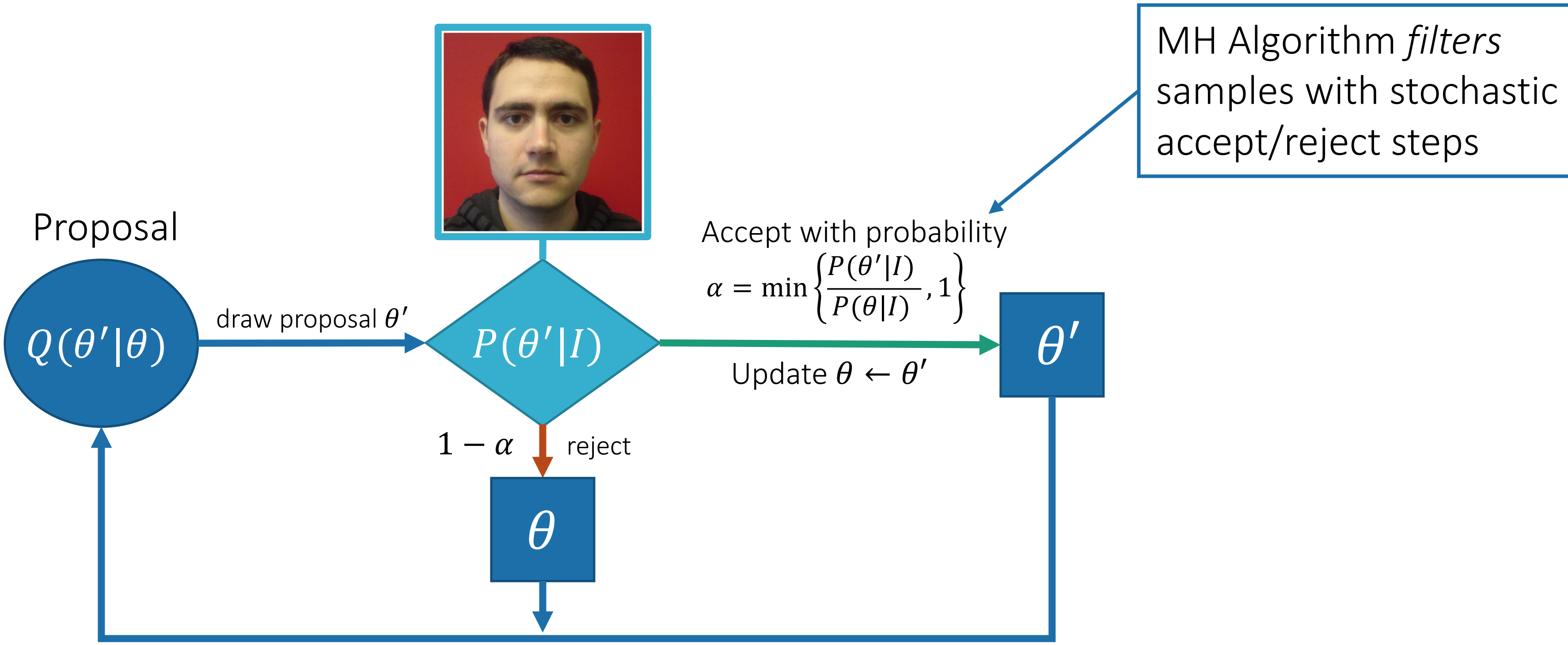
$$P\colon\ \tilde{P}(\theta|I) = \ell(\theta; I)P(\theta)$$

  - Unnormalized
  - Point-wise evaluation only

- Parameters
  - Shape: 50 – 200, low-rank parameterized GP shape model
  - Color: 50 – 200, low-rank parameterized GP color model
  - Pose/Camera: 9 parameters, pin-hole camera model
  - Illumination: 9*3 Spherical Harmonics illumination/reflectance

≈ 300 dimensions (!!)

Metropolis Algorithm



- Asymptotically generates samples $\theta_i \sim P(\theta|I)$: $\theta_1, \theta_2, \theta_3, \ldots$
- Markov chain Monte Carlo (MCMC) Method
- Works with *unnormalized*, point-wise posterior

Proposals

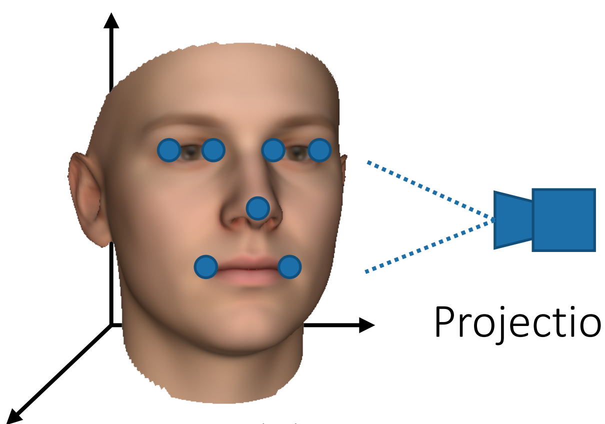
- Choose simple Gaussian random walk proposals (Metropolis)
$$"Q(\theta'|\theta) = N(\theta'|\theta, \Sigma_\theta)"$$
  - Normal *perturbations* of current state
- Block-wise to account for different parameter types
  - Shape $N(\boldsymbol{\alpha}'|\boldsymbol{\alpha}, \sigma_S^2 E_S)$
  - Color $N(\boldsymbol{\beta}'|\boldsymbol{\beta}, \sigma_C^2 E_C)$
  - Camera $\sum_c N(\theta_c'|\theta_c, \sigma_c^2)$
  - Illumination $\sum_i N(\theta_L'|\theta_L, \sigma_{L,i}^2 E_L)$

- Large mixture distributions, e.g.
$$\frac{2}{3} Q_P(\theta'|\theta) + \frac{1}{3} \sum_i \lambda_i Q_i^L(\theta'|\theta)$$

In practice, we often add more complicated proposals, e.g. shape scaling, a direct illumination estimation and decorrelation

Landmarks Fitting

Face Model

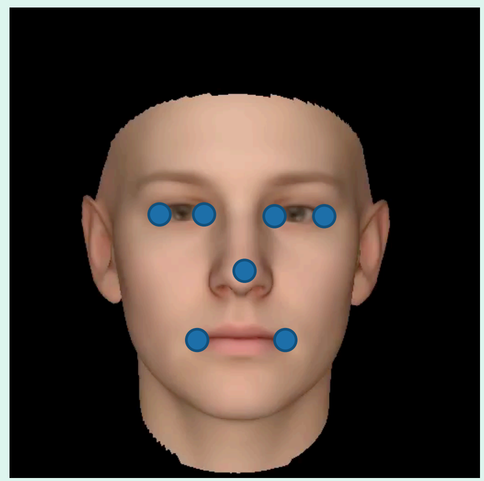


Projection

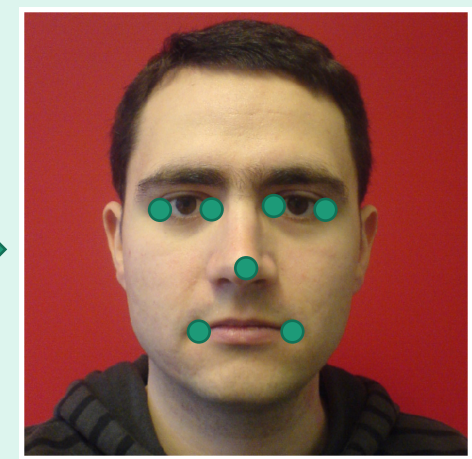
Prior $P(\theta)$

Variable Parameters
- Pose
- Shape

Rendered Landmarks



Target Landmarks



Likelihood $\ell(\theta; \tilde{\boldsymbol{x}}) \propto P(\tilde{\boldsymbol{x}} | \boldsymbol{x}(\theta))$

3DMM Landmarks Likelihood

Simple models: **Independent Gaussians**

- Observation of landmark locations in image
  - Single landmark position model:

$$\boldsymbol{x}_i^{\mathrm{2D}}(\theta) = (\mathrm{T_{VP}} \circ \mathrm{Pr} \circ \mathrm{T_{MV}} \circ h_\alpha)\left(\boldsymbol{x}_i^{\mathrm{3D}}\right)$$

$$\ell_i\left(\theta; \widetilde{\boldsymbol{x}}_i^{\mathrm{2D}}\right) = N\left(\widetilde{\boldsymbol{x}}_i^{\mathrm{2D}} \middle| \boldsymbol{x}_i^{\mathrm{2D}}(\theta), \sigma_{\mathrm{LM}}^2\right)$$

$$\mathrm{T_{MV}}(\boldsymbol{x}) = R_{\varphi,\psi,\vartheta}(\boldsymbol{x}) + \boldsymbol{t}$$

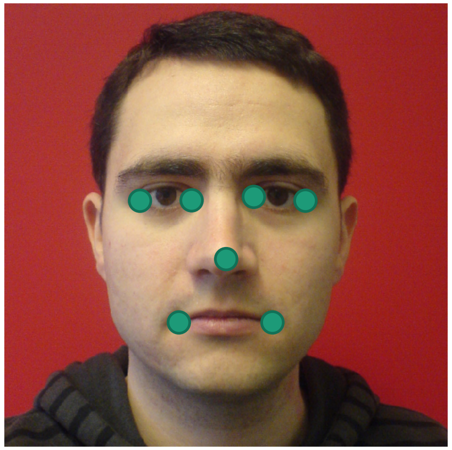
$$(\mathrm{T_{VP}} \circ \mathrm{Pr})(\boldsymbol{x}) = \begin{bmatrix} \frac{w}{2} * \frac{x}{z} \\ -\frac{h}{2} * \frac{y}{z} \end{bmatrix} + \boldsymbol{t}_{pp}$$

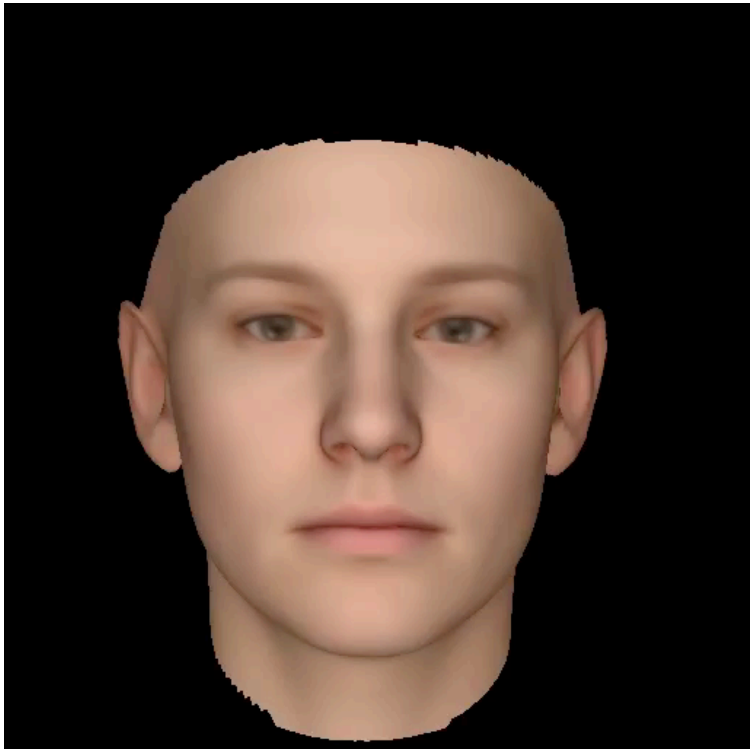
  - Independent model

$$\ell\left(\theta;\ \{\widetilde{\boldsymbol{x}}_i^{\mathrm{2D}}\}_i\right) = \prod_i \ell\left(\theta; \widetilde{\boldsymbol{x}}_i^{\mathrm{2D}}\right)$$

- Independence and Gaussian are just *simple models* (questionable)

Landmarks: Samples





Results: Landmarks

- Landmarks posterior:

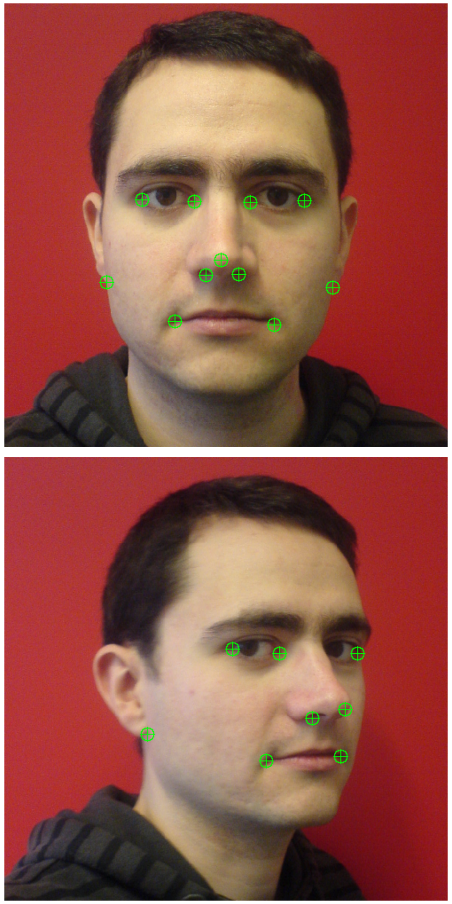
  Manual labelling: $\sigma_{\mathrm{LM}} = 4\mathrm{pix}$

  Image: 512x512

- Certainty of pose fit
  - Influence of ear points?
  - Frontal better than sideview?

Yaw, $\sigma_{\mathrm{LM}} = 4\mathrm{pix}$	with ears	w/o ears
Frontal	$1.4° \pm \mathbf{0.9}°$	$-1.4° \pm \mathbf{2.7}°$
Sideview	$24.8° \pm \mathbf{2.5}°$	$25.2° \pm \mathbf{4.0}°$

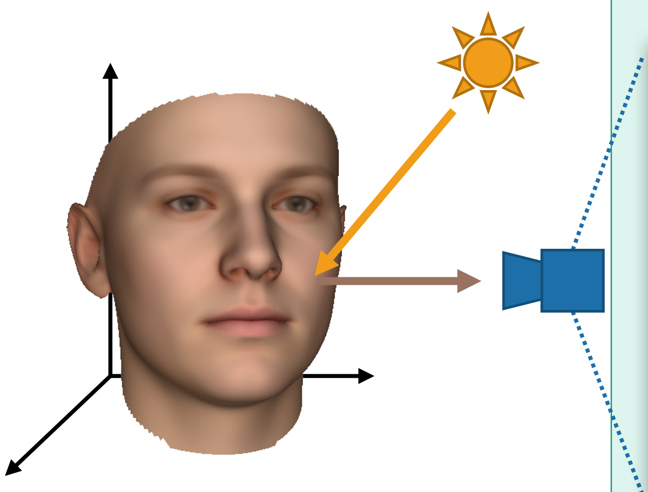
Distance stdev	with ears	w/o ears
Frontal	22cm	125cm
Sideview	35cm	35cm



Face Model Fitting

Reconstruction: Analysis-by-Synthesis

Face Model



Rendered Image $I(\theta)$

Target Image I

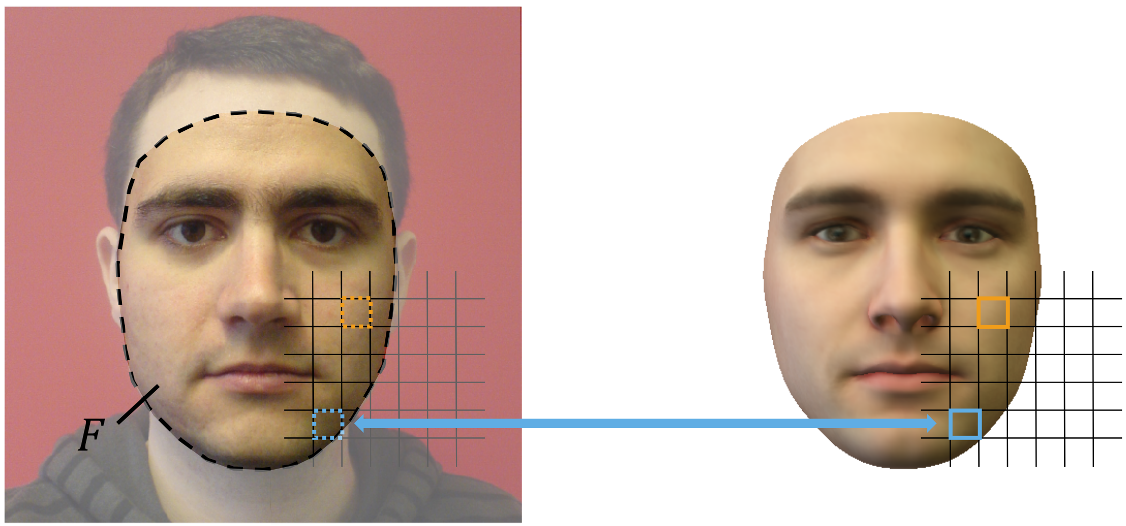
Parametric face model

Likelihood $\ell(\theta; I) \propto P(I|\, I(\theta))$

$\theta = (\vartheta, \alpha, \beta)$: ϑ Scene Parameters, α Face shape, β Face color

Independent Pixels Likelihood

Standard choice
Corresponds to least squares fitting



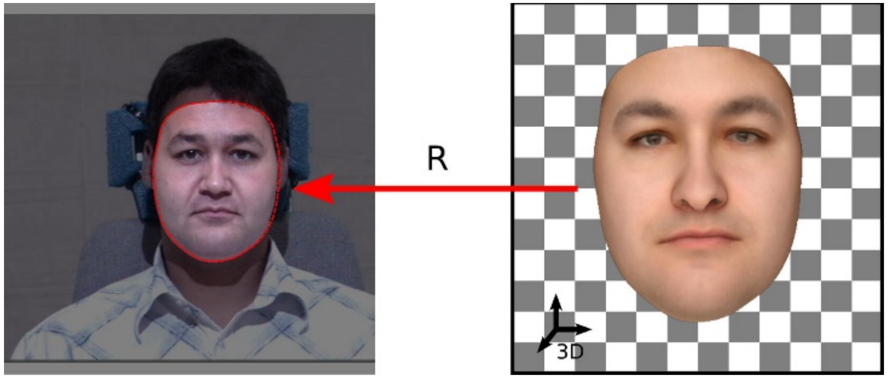
$$\ell(\theta;\tilde{I}) = \mathcal{N}(\blacksquare \mid \blacksquare, \sigma^2 I_3) * \mathcal{N}(\blacksquare \mid \blacksquare, \sigma^2 I_3) * \cdots$$

$$\ell(\theta;\tilde{I}) = \prod_{i \in F} \mathcal{N}\left(\tilde{I}_i \mid I_i(\theta), \sigma^2 I_3\right)$$

Background Model

The face model covers only a small part of the complete target image

$$\ell(\theta;\tilde{I}) = \prod_{i \in F} \ell_i(\theta;\tilde{I}_i)$$



What to do *outside* face region?

- *Ignore* → strong artifacts
- *Explicit* model

Shrinking   Misalignment

Explicit Background Model

Add explicit likelihood for background

$$\ell(\theta;\tilde{I}) = \prod_{i\in F} \ell_{\mathrm{F}}(\theta;\tilde{I}_i) \prod_{j\in B} b_{\mathrm{BG}}(\tilde{I}_i)$$

Why is ignoring bad?

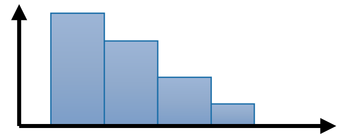
$$\ell(\theta;\tilde{I}) = \prod_{i\in F} \ell_{\mathrm{F}}(\theta;\tilde{I}_i) \Leftrightarrow b_{\mathrm{BG}}(\tilde{I}_i) = 1$$

Implicit background model is ***always*** present but might be inappropriate
→ better make it explicit!

Arbitrary background: The explicit background model needs to be based on *generic* and *simple* assumptions:

Constant model

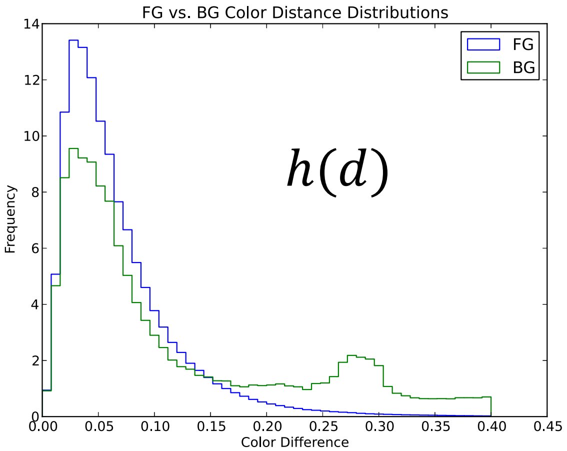
Histogram model



Schönborn et al. «Background modeling for generative image models», Computer Vision and Image Understanding, Volume 136, July 2015, Pages 117–127, doi:10.1016/j.cviu.2015.01.008

Collective Likelihood

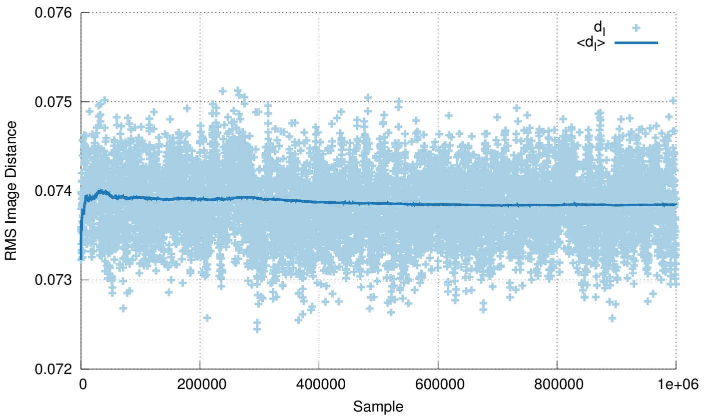
- Independence is not a good assumption
  - Too many observations (100k+): overconfident
  - Colors are correlated
- Model distribution of image distance
  - Fit to *empirical* histogram or use model
  - Can be any measure extracted on images
- Most-likely solutions match the image with the expected noise level
  - A perfect reconstruction is *unlikely*



Posterior Samples: Fitting Result

- Model instances with comparable reconstruction quality
- Remaining uncertainty of model representation
- Integration of uncertain detection directly into model adaptation





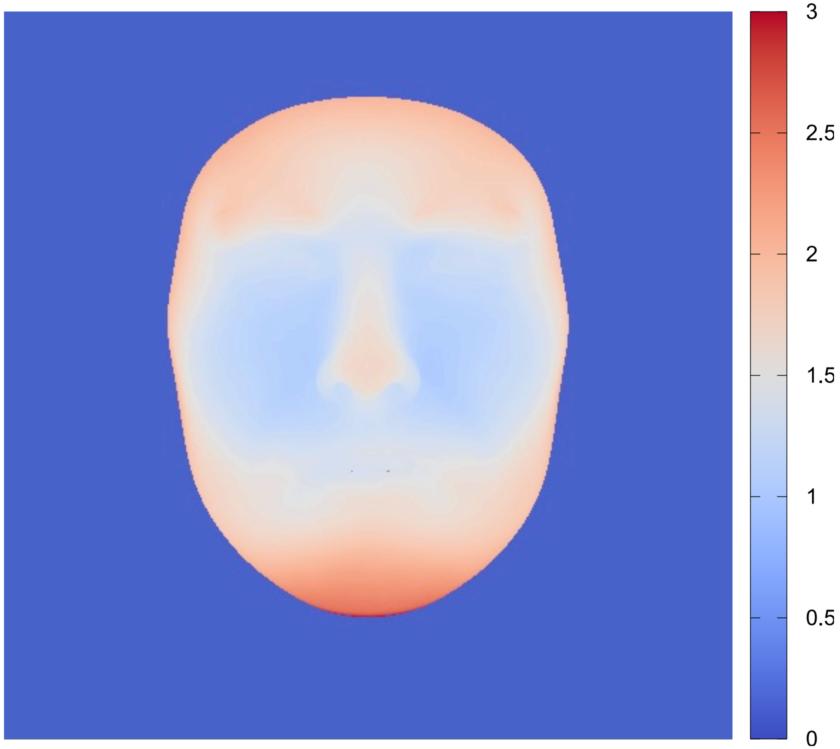
Posterior using *collective likelihood*

Results: Image

Yaw angle: $1.9° \pm 0.2°$

Image: Samples

Posterior Shape Variation



Landmarks posterior,
sd[mm]

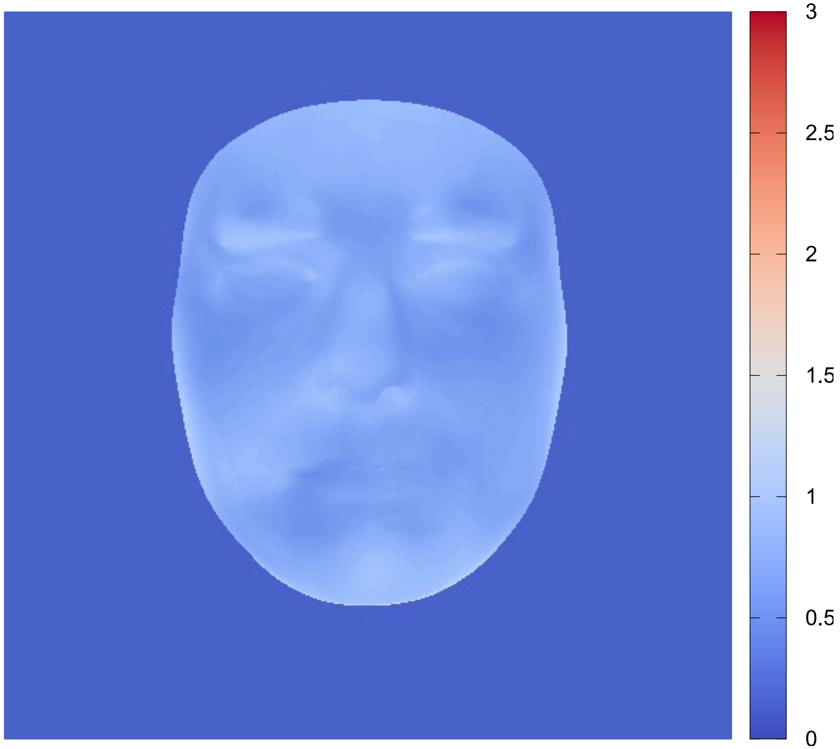
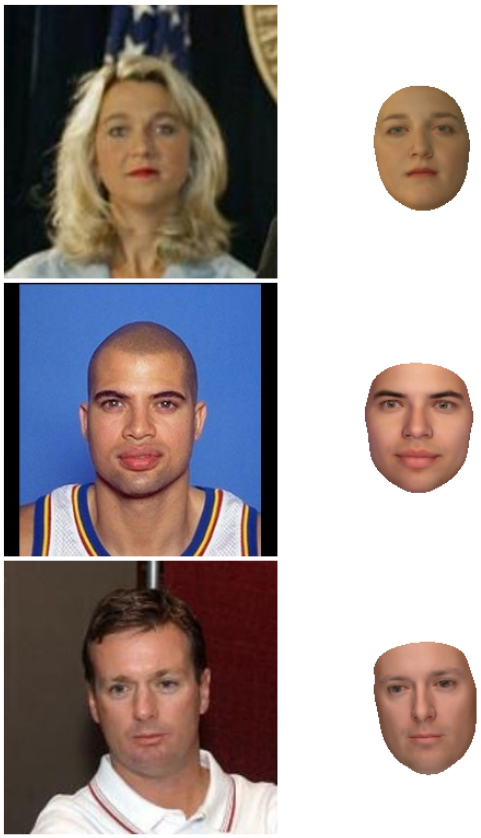


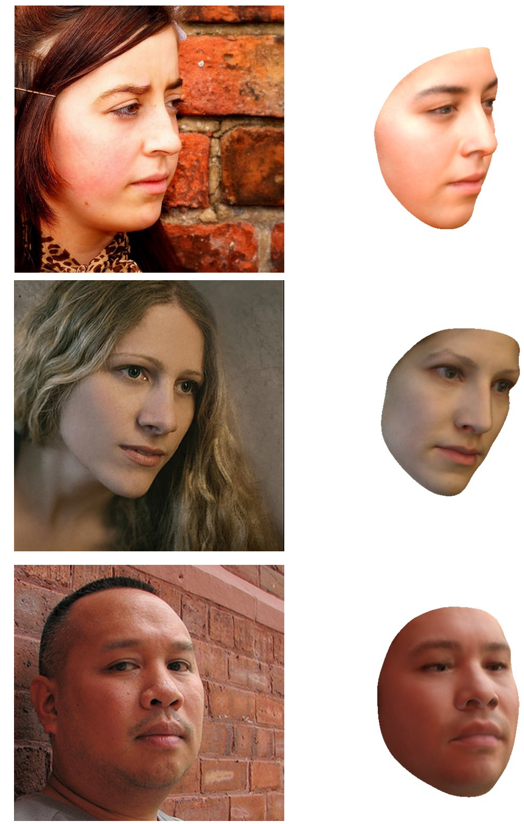
Image posterior,
sd[mm]

Fitting Results



LFW

Images from: Huang, Gary B., et al. *Labeled faces in the wild: A database for studying face recognition in unconstrained environments*. Vol. 1. No. 2. Technical Report 07-49, University of Massachusetts, Amherst, 2007.



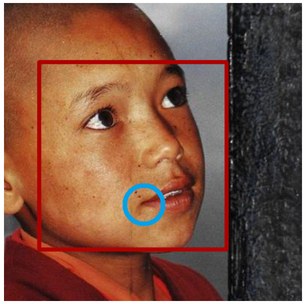
AFLW

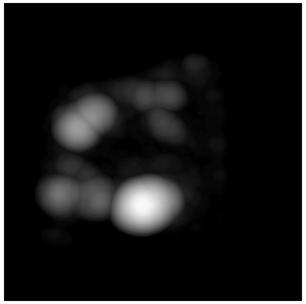
Images from: Köstinger, Martin, et al. "Annotated facial landmarks in the wild: A large-scale, real-world database for facial landmark localization." *Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on*. IEEE, 2011.

Automatic Fitting

- Detection of *face* and *feature points*
    - Scanning window & classifier
    - Uncertain results
    - Feed-forward: early *hard* decisions
- Integration concept
    - Bayesian integration
      → ***Filtering***
    - Metropolis sampling
      → ***Propose & verify***

Which box contains the face?

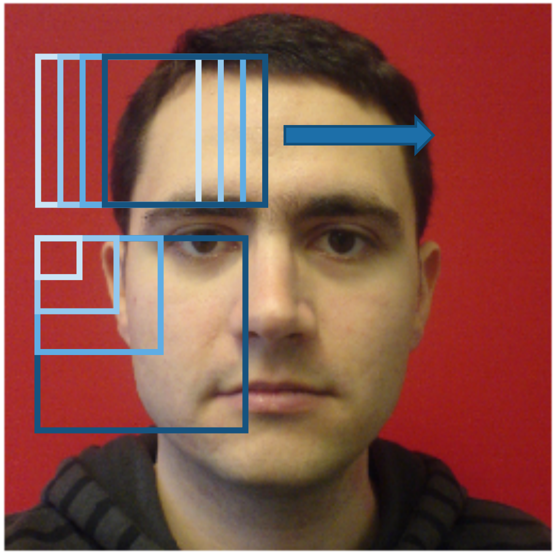




Schönborn, Sandro, et al. "Markov Chain Monte Carlo for Automated Face Image Analysis." *International Journal of Computer Vision* (2016): 1-24.

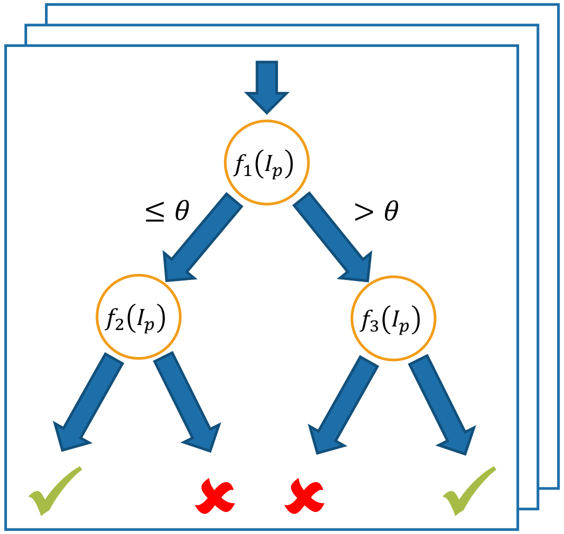
Random Forest Detection

- Scanning Window



- Classify each patch: face or not
- Search over image
- Search over scales

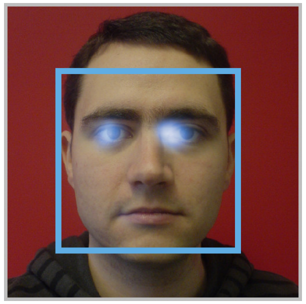
- Random Forest Classifier



- Haar Features 
- Information gain splitting
- Bagging many trees, depth ~16
- ~200k training patches (AFLW)

Bayesian Integration

Detection data



Bayesian integration

Observation likelihood

$$\ell(\theta; F, D) = P(F|\theta)P(D|\theta)$$

Bayesian inference

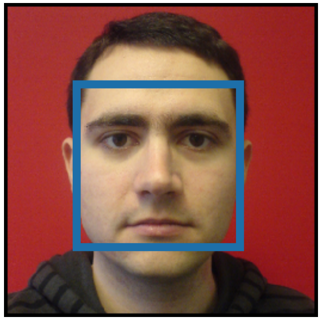
$$P(\theta|F, D) = \frac{\ell(\theta; F, D)P(\theta)}{N(F, D)}$$

- Different *modality*
  - Box F: position & size
  - Landmarks D: certainty
- Detection is uncertain

- *Likelihood* models
  - Detection is *observation*
  - Different observation models
- Conceptual uncertainty

Detection Likelihood

Face detection

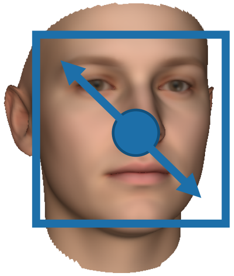


Box: position & size of detected face

$$\boldsymbol{p}, s$$

Model: Uncertainty of position and scale

$$\ell(\theta; F) = \mathcal{N}\left(\boldsymbol{p} | \boldsymbol{x}(\theta), \sigma_{\mathrm{F}}^2\right) \mathcal{LN}\left(s | s(\theta), \sigma_{\mathrm{S}}^2\right)$$



Landmarks detection

Detection map: *certainty* of detecting at position $\boldsymbol{x}$

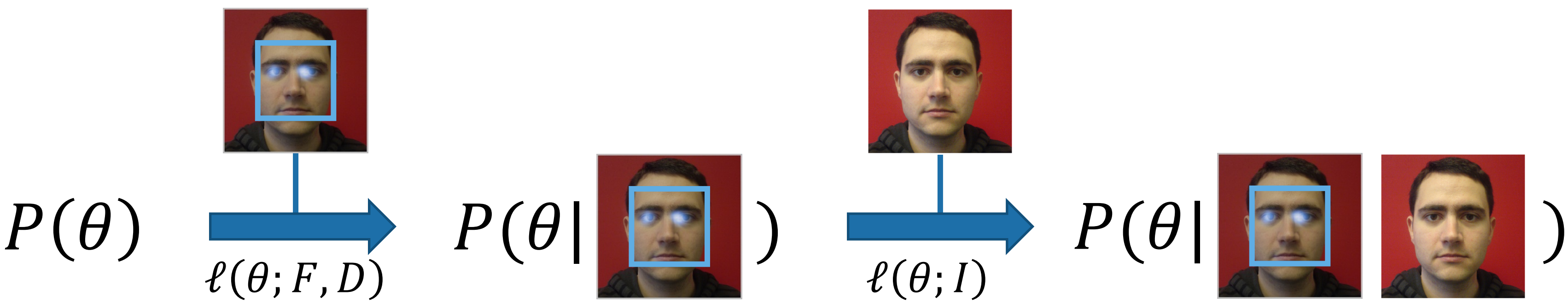
$$D(\boldsymbol{x})$$

Model: Best combination of landmarks uncertainty and detection certainty

$$\ell(\theta; D) = \max_{\boldsymbol{t}} \mathcal{N}(\boldsymbol{t} | \boldsymbol{x}(\theta), \sigma^2) D(\boldsymbol{t})$$

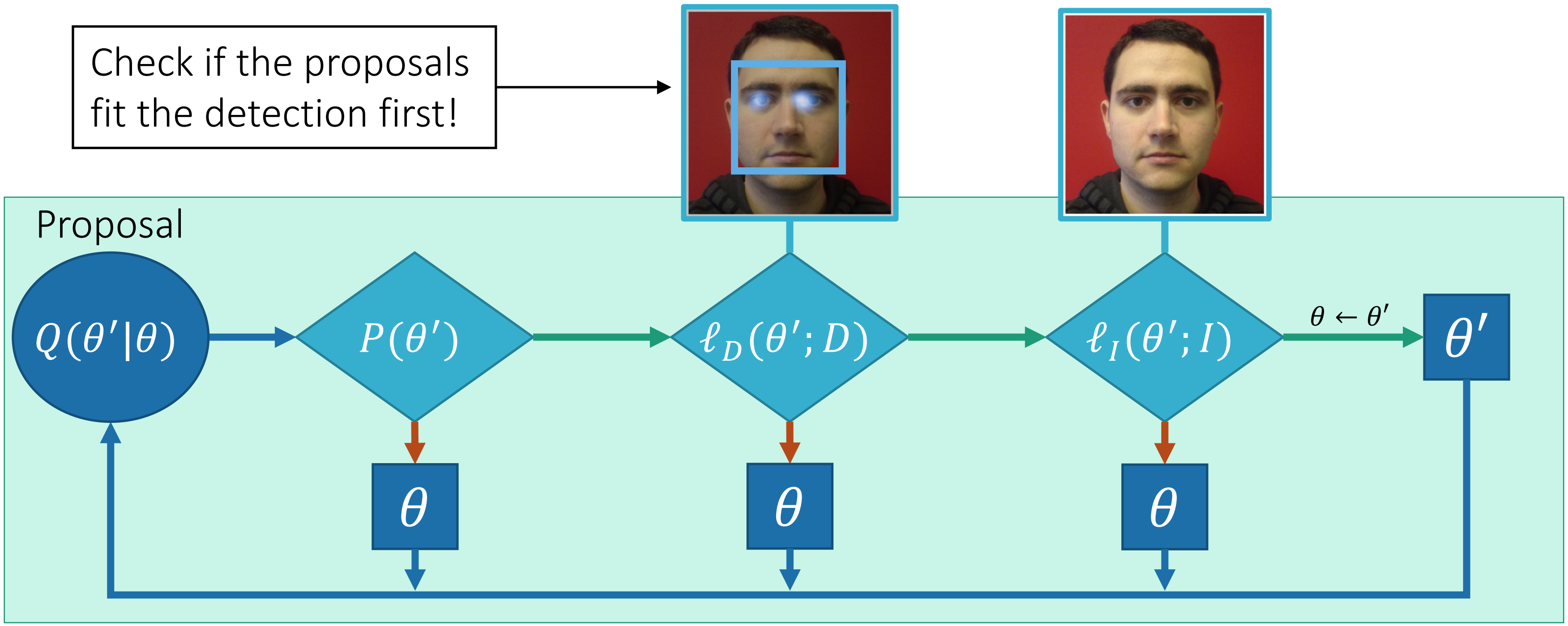
Integration by Filtering

- Step-by-step Bayesian inference



- Condition on observations *one after the other*
- *Posterior* of first observation becomes *prior* for next step
  - Each step adds an observation through conditioning with its likelihood
- Equivalent to single-step Bayesian inference

Filtering: Multiple Metropolis Decisions

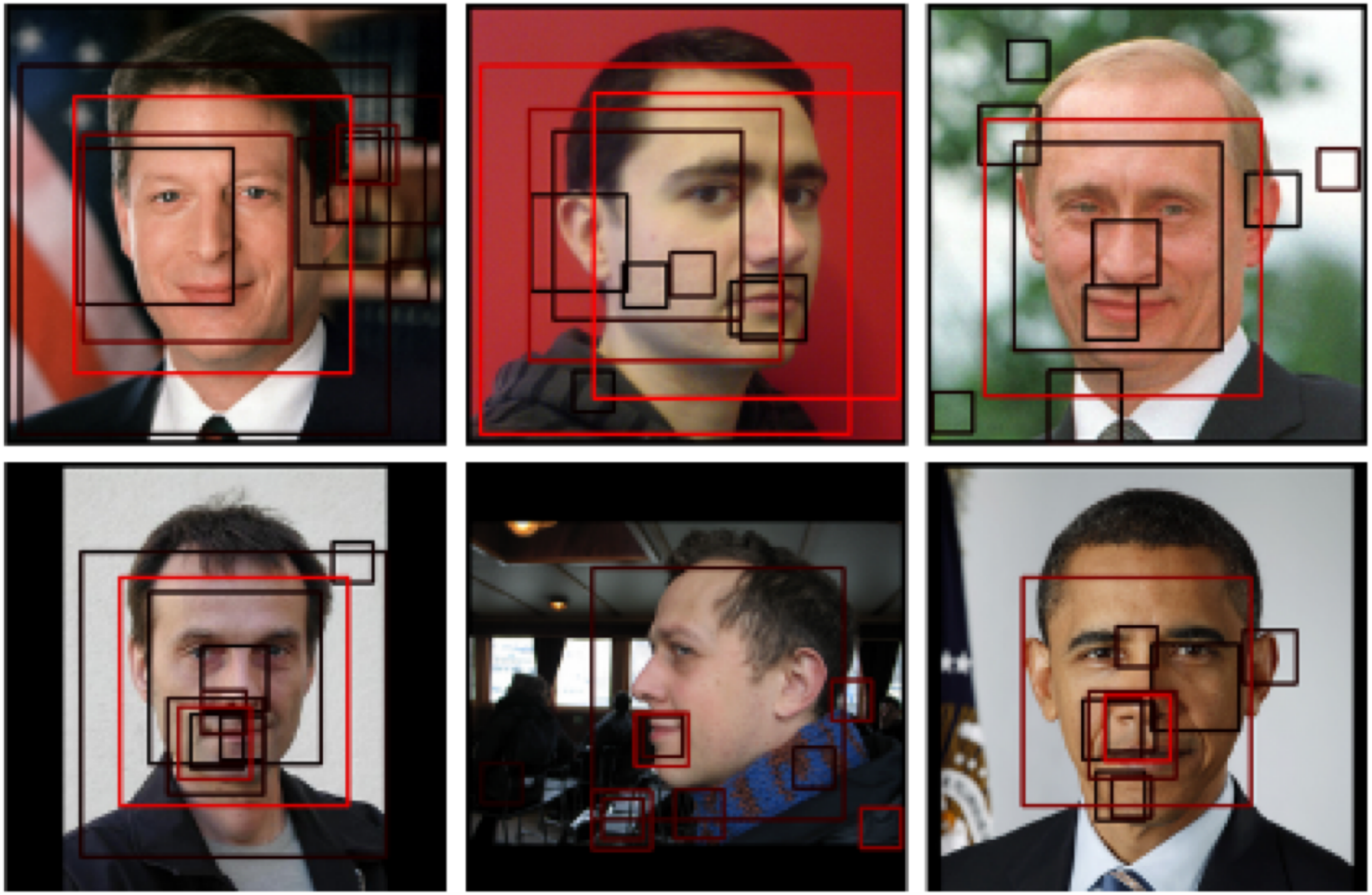


$$Q \xrightarrow{P} P(\theta) \xrightarrow{F,D} P(\theta|F,D) \xrightarrow{I} P(\theta|F,D,I)$$

Bayesian inference steps

- Step-wise Bayesian inference: Needs $\ell(\theta)$ for each step
- Saves computation time if properly ordered

Alternatives



Metropolis: Propose-and-Verify

- Metropolis algorithm formalizes: *propose-and-verify*

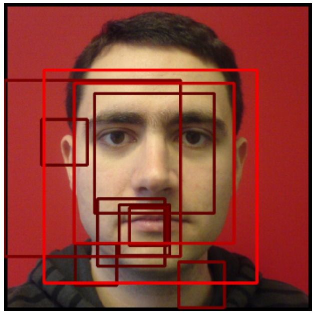
Draw a sample x' from $Q(x'\|x)$	Propose
With *probability* $\alpha = \min\left\{\frac{P(\boldsymbol{x}')}{P(\boldsymbol{x})}, 1\right\}$ accept $\boldsymbol{x}'$ as new sample	Verify

- Decouples *finding* possible solution from *selection*
  - No need to always provide good solutions in proposals
  - Verification for consistency with the model
- Algorithmic advantage beyond probabilistic Bayesian concept

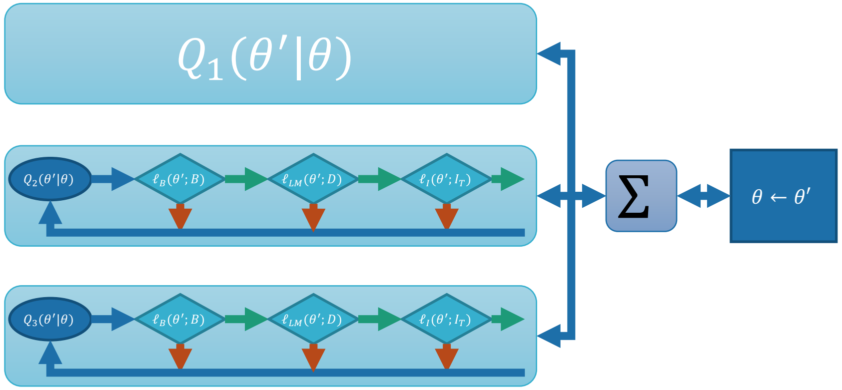
"Anything that is more informed *than*
random walks should improve fitting"

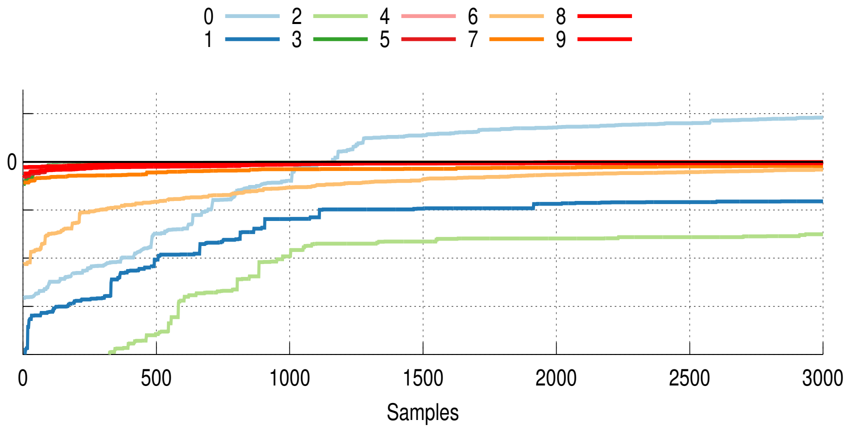
Multiple Alternative Proposals

- Metropolis formalizes *propose-and-verify*
  - Decouples *proposing* possible solution from *validation*
  - No need to always provide good solutions in proposals
- Introduce alternatives through *proposals*



Many candidates





Data-Driven Markov Chain Monte Carlo (DDMCMC):
Use *data* to build more *informed* proposals

"Anything that is more informed *than random walks should improve fitting"*

by courtesy of keystone

Summary

- Fitting as *probabilistic inference*
- Probabilistic inference is often intractable
- Sampling methods *approximate* by *simulation*
- MCMC methods provide a powerful sampling framework
  - Markov Chain with target distribution as equilibrium distribution
  - General algorithms, e.g. Metropolis-Hastings
- Fitting of the 3DMM as a real inference problem
- MH algorithm to integrate information: Framework
  - *Filtering:* Uncertain information as observation, step-by-step
  - *Propose-and-verify:* Alternatives, multiple hypotheses, heuristics