graphics and vision gravis University of Basel

2D Face Image Analysis

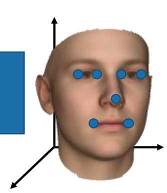
Probabilistic Morphable Model Fitting

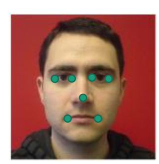
Basel2018

University of Basel

Contents

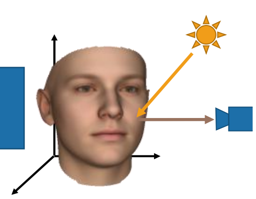
Landmarks Fitting

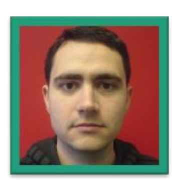




Observed Landmarks in 2D

Image Fitting



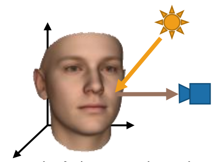


Observed Image

UNIVERSITÄT BASEL

> DEPARTMENT OF MATHEMATICS AND COMPUTER SCIENCE GRAVIS2018 ¦ BASEL

2D Face Image Analysis



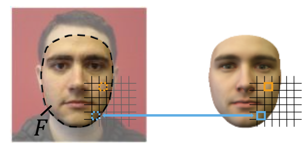
Morphable Model adaptation to explain image

Bayesian Inference Setup

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

Image Likelihood

Image as observation



Face & Feature point detection

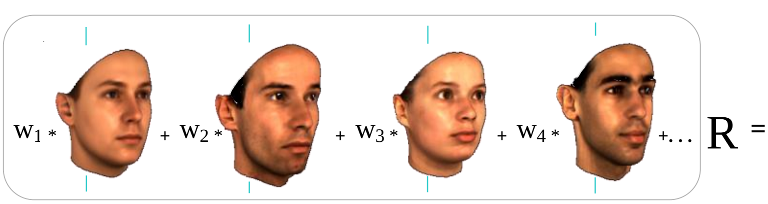
Integration of fast bottom-up methods

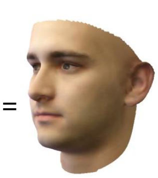
UNIVERSITÄT BASEL

> DEPARTMENT OF MATHEMATICS AND COMPUTER SCIENCE GRAVIS2018 ¦ BASEL

Computer Graphics: Rendering Faces

3D Face Scans	2D Images





Faces: GP models for shape & color:

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

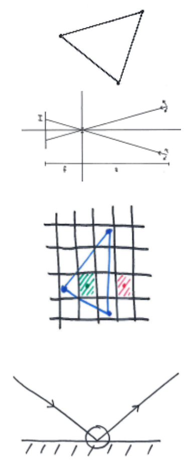
$$c_\beta = \mu + UD\beta \qquad \beta \sim N(0, I_d)$$

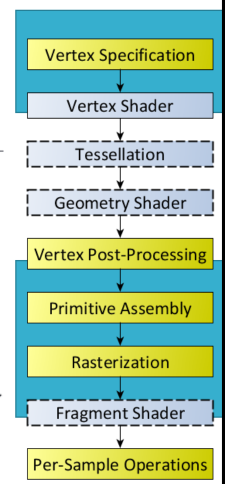
UNIVERSITÄT BASEL

> DEPARTMENT OF MATHEMATICS AND COMPUTER SCIENCE GRAVIS2018 ¦ BASEL

Computer Graphics Overview

- **Geometry** (result of shape modelling)
- **Camera & Projection**
  Transformations in space and projection
  Maps 3D space and 2D image plane
- **Rasterization**
  Correspondence: image pixels ↔ surface
  Z-Buffer: Hidden surface removal
- **Shading**
  Illumination simulation models
  - **Illumination**
    Phong: Ambient, diffuse & specular
    Global Illumination





5

UNIVERSITÄT BASEL

> DEPARTMENT OF MATHEMATICS AND COMPUTER SCIENCE GRAVIS2018 ¦ BASEL

Face-to-Image Transformations

- **Model-View**

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

- **Projection**

$$\mathcal{P}(x) = \frac{f}{z}\begin{bmatrix} x \\ y \end{bmatrix}$$

- **Viewport**

$$T_{VP}(x) = \begin{bmatrix} \frac{w}{2}(x+1) \\ \frac{h}{2}(1-y) \end{bmatrix} + \boldsymbol{t}_{pp}$$

- 9 Parameters:
  - (3) Translation $\boldsymbol{t}$
  - (3) Rotation φ, ψ, ϑ
  - (1) Focal length f
  - (2) Image Offset $\boldsymbol{t}_{pp}$
- 2 Constants:
  - (2) Image size / sampling

6

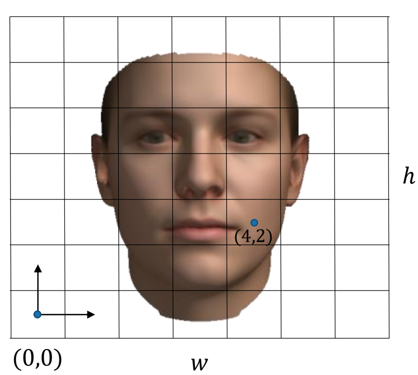
Perspective Effect

- Perspective division distorts image non-linearly
- Effect depends on relation of object depth and camera distance

UNIVERSITÄT BASEL
> DEPARTMENT OF MATHEMATICS AND COMPUTER SCIENCE GRAVIS2018 ¦ BASEL

Rasterization

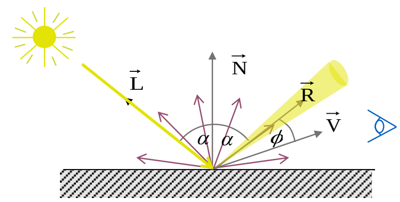
- Camera: 3D → 2D transformation for *points*
- Raster Image in image plane
- Establishes correspondence to 3D surface for each *pixel*
- Basis: geometric *primitives*



Pixel grid, cell-centered

Phong Illumination Model

- Combination of three illumination contributions:
  - Lambert (diffuse) $k_{\text{diff}} * I_L * \cos(L, N)$
  - Specular $k_{\text{spec}} * I_L * \cos(\mathrm{R}, \mathrm{V})^n$
  - Ambient (global) $k_{\text{amb}} * I_A$



- Ambient is a scene *average* light intensity I_A
- Lambert and specular part for each light source

$$I' = k_{\text{amb}} * I_A + k_{\text{diff}} * I_L * \cos(L, N) + k_{\text{spec}} * I_L * \cos(\mathrm{R}, \mathrm{V})^n$$

usually colored

UNIVERSITÄT BASEL
> DEPARTMENT OF MATHEMATICS AND COMPUTER SCIENCE GRAVIS2018 ¦ BASEL

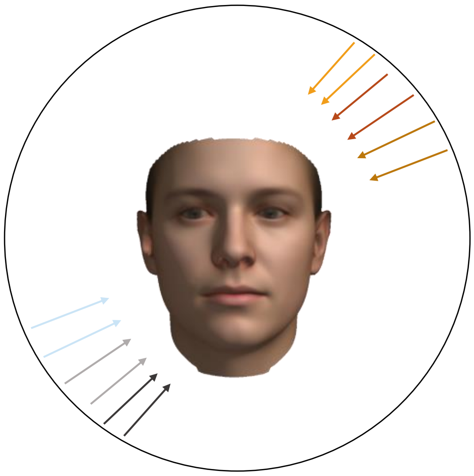
Phong Illumination Model

- Combination of three illumination contributions:
  - Lambert (diffuse) $k_{\text{diff}} * I_L * \cos(L, N)$
  - Specular $k_{\text{spec}} * I_L * \cos(\mathrm{R}, \mathrm{V})^n$
  - Ambient (global) $k_{\text{amb}} * I_A$
- Ambient is a scene *average* light intensity I_A
- Lambert and specular part for each light source

$$I' = k_{\text{amb}} * I_A + k_{\text{diff}} * I_L * \cos(L, N) + k_{\text{spec}} * I_L * \cos(\mathrm{R}, \mathrm{V})^n$$

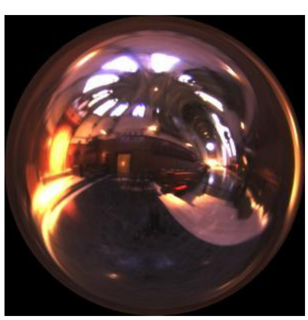
usually colored

Environment Maps



- Mapping of incoming light intensity from every direction
  $$I_L^{\mathrm{RGB}}(\theta, \varphi)$$
- Modeled at infinity
- Typically *empirically* captured
- Shading with environment maps requires *integration* over all incoming directions

Environment Maps



Grace Cathedral (San Francisco)
P. Debevec

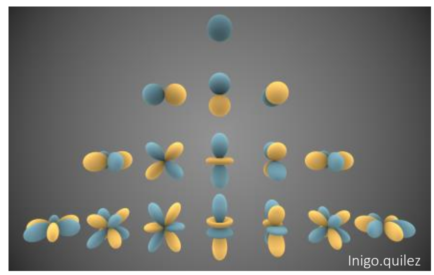
White surface in Grace Cathedral

UNIVERSITÄT BASEL

> DEPARTMENT OF MATHEMATICS AND COMPUTER SCIENCE GRAVIS2018 ¦ BASEL

Spherical Harmonics Illumination

- Expand map $I_L^{\mathrm{RGB}}(\theta, \varphi)$ with *basis* functions
- Choose *Spherical Harmonics*: Eigenfunctions of Laplace operator on sphere surface
  $Y_{lm}(\theta, \varphi)$
- Corresponds to Fourier transform
- Integration becomes multiplication of coefficients (→ *fast convolution*)
- Low frequency part is sufficient for Lambertian reflectance



Ramamoorthi, Ravi, and Pat Hanrahan. "An efficient representation for irradiance environment maps." Proceedings of the 28th annual conference on Computer graphics and interactive techniques. ACM, 2001.

13

UNIVERSITÄT BASEL

> DEPARTMENT OF MATHEMATICS AND COMPUTER SCIENCE GRAVIS2018 ¦ BASEL

Environment Map Illumination

14

UNIVERSITÄT BASEL

> DEPARTMENT OF MATHEMATICS AND COMPUTER SCIENCE GRAVIS2018 ¦ BASEL

Image Formation: at each Vertex k

α_i ↓ x, y, z ↓ Rigid Transformation ↓ Perspective Projection ↓ p_x, p_y

β_i ↓ r, g, b ↓ Illumination Model ↓ Color Transformation ↓ I_r, I_g, I_b

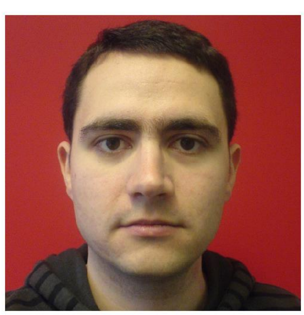
Rigid Transformation → Normals → Illumination Model

$$\mathbf{I}_{\text{model}}(p_x, p_y) = I_r, I_g, I_b$$

UNIVERSITÄT BASEL

> DEPARTMENT OF MATHEMATICS AND COMPUTER SCIENCE GRAVIS2018 ¦ BASEL

3D Face Reconstruction



16

Probabilistic Inference for Image Registration

- Generative image explanation: How to find θ explaining I ?

$$p(\theta|I) = \frac{\ell(\theta;I)\,p(\theta)}{N(I)} \qquad N(I) = \int \ell(\theta;I)p(\theta)\mathrm{d}\theta$$

-----> Normalization intractable in our setting

- What can be done:
  1. Accept MAP as the only option
  2. Approximate posterior distribution (e.g. use sampling methods)

MH Inference of the 3DMM

- Target distribution is our "**posterior**":

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

  - Unnormalized
  - Point-wise evaluation only

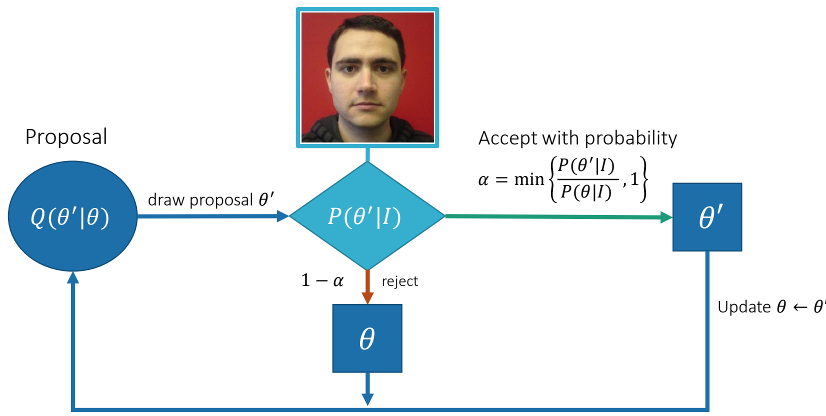
- **Parameters**
  - Shape: 50 – 200, low-rank parameterized GP shape model
  - Color: 50 – 200, low-rank parameterized GP color model
  - Pose/Camera: 9 parameters, pin-hole camera model
  - Illumination: 9*3 Spherical Harmonics illumination/reflectance

$\approx$ 300 dimensions (!!)

UNIVERSITÄT BASEL

> DEPARTMENT OF MATHEMATICS AND COMPUTER SCIENCE GRAVIS2018 ¦ BASEL

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

21

UNIVERSITÄT BASEL

> DEPARTMENT OF MATHEMATICS AND COMPUTER SCIENCE GRAVIS2018 ¦ BASEL

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 $\sum_i N(\theta_L'|\theta_L, \sigma_{L,i}^2 E_L)$

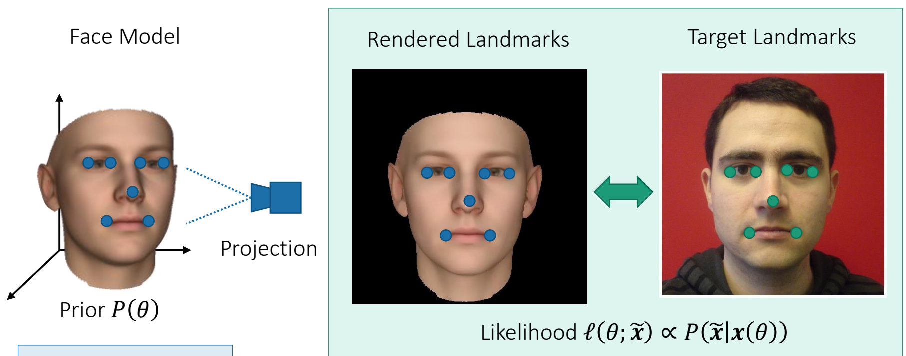
In practice, we often add more complicated proposals, e.g. shape scaling, a direct illumination estimation and decorrelation

- Large mixture distributions, e.g.

$$\frac{2}{3} Q_P(\theta'|\theta) + \frac{1}{3} \sum_i \lambda_i Q_i^L(\theta'|\theta)$$

22

Landmarks Fitting



Variable Parameters
- Pose
- Shape

3DMM Landmarks Likelihood

Simple models: **Independent Gaussians**

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

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

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

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

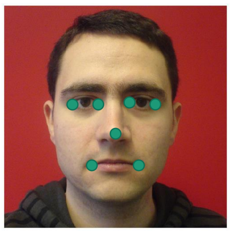
$$(\mathrm{T}_{\mathrm{VP}} \circ \mathrm{Pr})(\boldsymbol{x}) = \begin{bmatrix} \frac{w}{2} * \frac{x}{z} \\ -\frac{h}{2} * \frac{y}{z} \end{bmatrix} + \boldsymbol{t}_{pp}$$

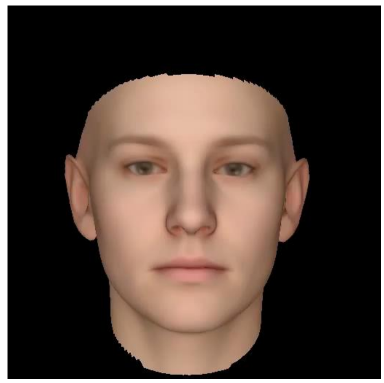
  - Independent model

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

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

Landmarks: Samples





Results: 2D Landmarks

- Landmarks posterior:

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

  Image: 512x512

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

Yaw, $\sigma_{\text{LM}} = 4\text{pix}$	with ears	w/o ears
Frontal	$1.4^\circ \pm \mathbf{0.9}^\circ$	$-0.8^\circ \pm \mathbf{2.7}^\circ$
Side view	$24.8^\circ \pm \mathbf{2.5}^\circ$	$25.2^\circ \pm \mathbf{4.0}^\circ$

- Landmarks posterior:

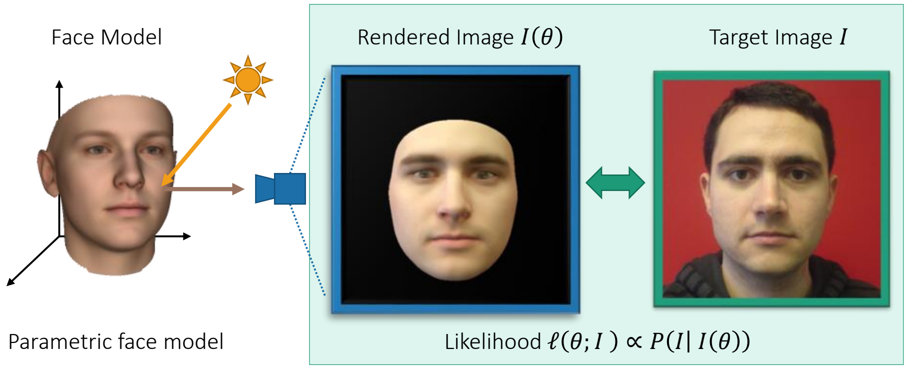
  Manual labelling: $\sigma_{\text{LM}} = 4\text{pix}$

  Image: 512x512

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

Face Model Fitting

Reconstruction: Analysis-by-Synthesis

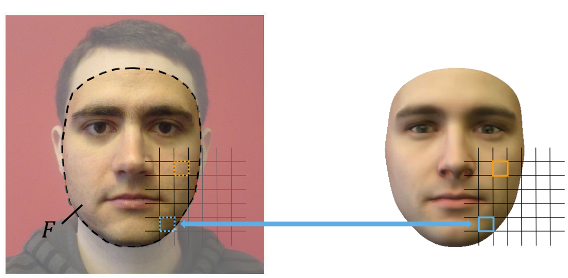


$\theta = (\vartheta, \alpha, \beta)$: ϑ Scene Parameters, α Face shape, β Face color

UNIVERSITÄT BASEL
> DEPARTMENT OF MATHEMATICS AND COMPUTER SCIENCE GRAVIS2018 ¦ BASEL

Independent Pixels Likelihood

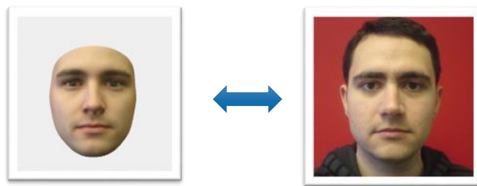
Standard choice
Corresponds to least squares fitting



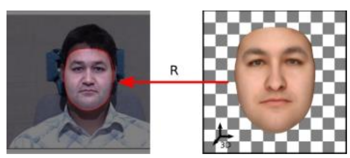
$$\ell(\theta; \tilde{I}) = \mathcal{N}(\blacksquare \mid \blacksquare, \sigma^2 I_3) * \mathcal{N}(\blacksquare \mid \blacksquare, \sigma^2 I_3) * \cdots$$

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

Image Likelihood

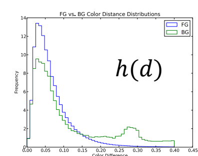


Background model is required
The face model does not cover the complete target image and shows self-occlusion.



Collective likelihood model
Pixels are not independent. We can also model the empirical distribution of image distance

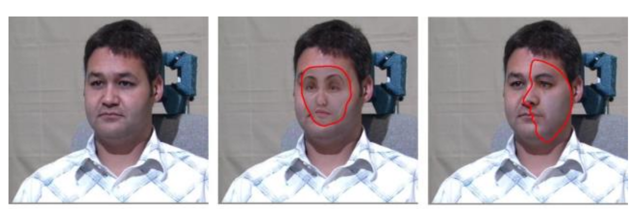
$d = \|$  $-$  $\|$



Background Model

Schönborn et al. 2015
«Background modeling for generative image models»,
Computer Vision and Image Understanding, Volume 136

Face covers only parts of the image – background must not be ignored

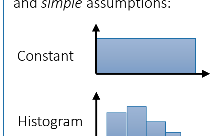


(a) Target (b) Shrinking (c) Pose Misalignment

- Variable alignment of model with the image
  - Projected size and self-occlusion
  - Shrinking or misalignment
- Model background pixels explicitly

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

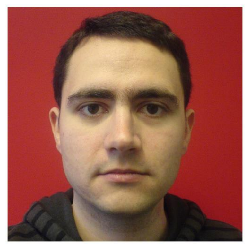
Arbitrary background: The explicit background model needs to be based on *generic* and *simple* assumptions:

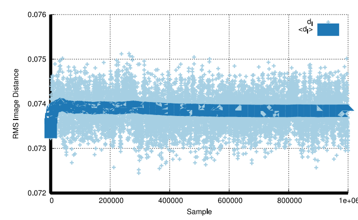


32

Posterior Samples: Fitting Result

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

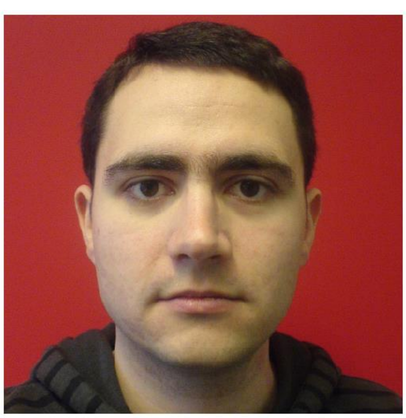




Posterior using *collective likelihood*

UNIVERSITÄT BASEL
> DEPARTMENT OF MATHEMATICS AND COMPUTER SCIENCE GRAVIS2018 ¦ BASEL

Results: Image



Yaw angle: $1.9° \pm 0.2°$

UNIVERSITÄT BASEL

> DEPARTMENT OF MATHEMATICS AND COMPUTER SCIENCE GRAVIS2018 ¦ BASEL

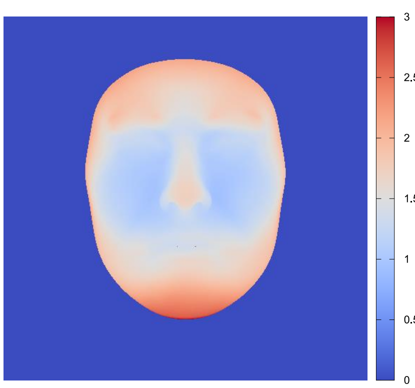
Image: Samples

35

UNIVERSITÄT BASEL

> DEPARTMENT OF MATHEMATICS AND COMPUTER SCIENCE GRAVIS2018 ¦ BASEL

Posterior Shape Variation



Landmarks posterior,
sd[mm]

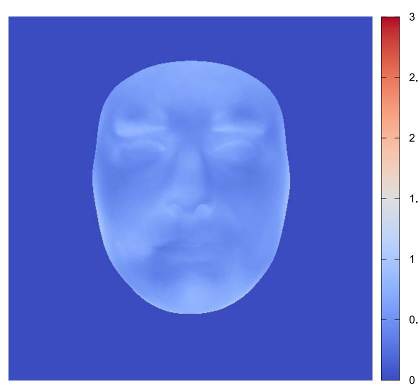


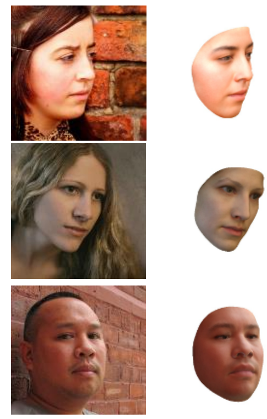
Image posterior,
sd[mm]

36

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

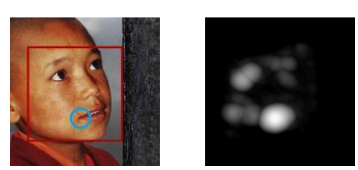
Images from: Köstinger, Martin, et al. "Annotated facial landmarks in the wild: A large-scale, real-world database for facial landmark localization." *Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on*. IEEE, 2011.

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***

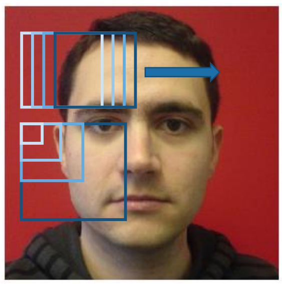
Which box contains the face?



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

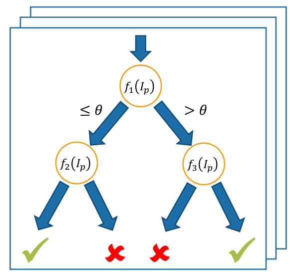
Random Forest Detection

- Scanning Window



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

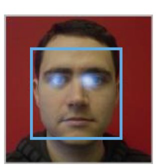
- Random Forest Classifier



- 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

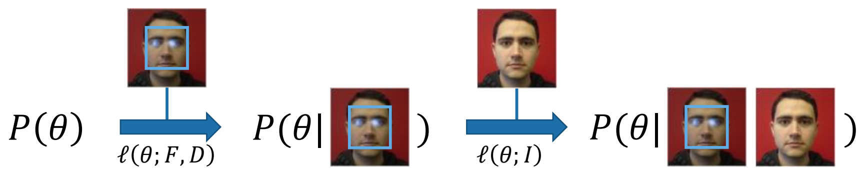
$$P(\theta|F, D) = \frac{\ell(\theta; F, D)P(\theta)}{N(F, D)}$$

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

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

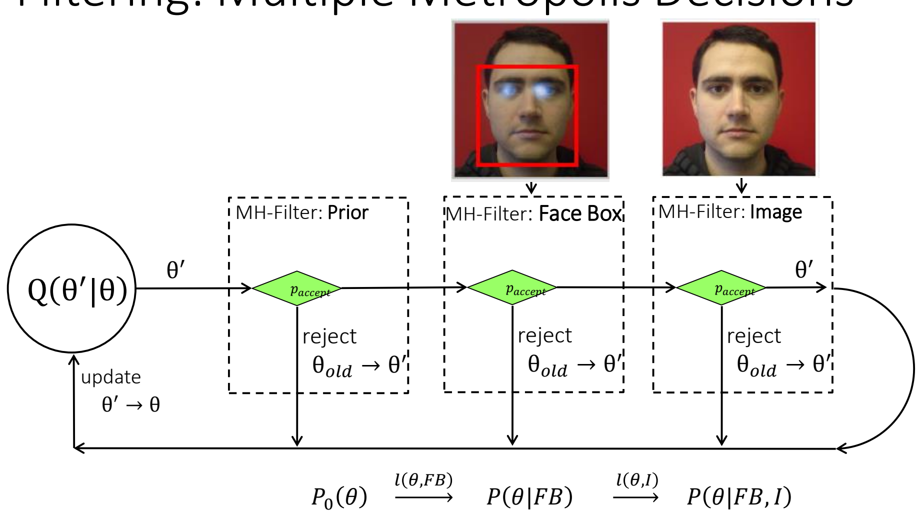
Integration by Filtering

- 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



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

UNIVERSITÄT BASEL

> DEPARTMENT OF MATHEMATICS AND COMPUTER SCIENCE GRAVIS2018 ¦ BASEL

by courtesy of keystone

UNIVERSITÄT BASEL

> DEPARTMENT OF MATHEMATICS AND COMPUTER SCIENCE GRAVIS2018 ¦ BASEL

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

Occlusion-aware 3D Morphable Face Models

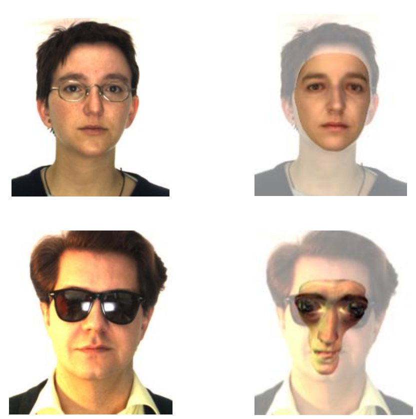
Bernhard Egger, Sandro Schönborn, Andreas Schneider, Adam Kortylewski, Andreas Morel-Forster, Clemens Blumer and Thomas Vetter

International Journal of Computer Vision, 2018

UNIVERSITÄT BASEL
> DEPARTMENT OF MATHEMATICS AND COMPUTER SCIENCE GRAVIS2018 ¦ BASEL

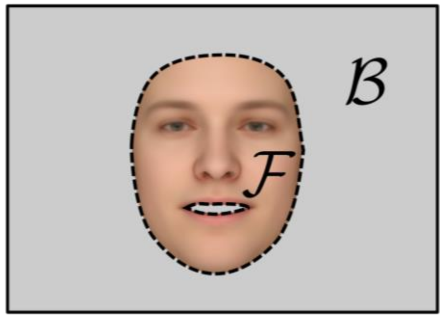
Face Image Analysis under Occlusion

Source: AFLW Database



Source: AR Face Database

There is nothing like: no background model

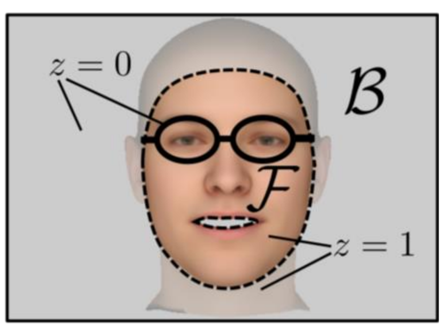


$$\ell(\theta;I) = \prod_{x\in I} \ell(\theta;I(x)) = \prod_{i\in F} l_{face}(\theta;\tilde{I}_i) \prod_{i`\in B} b(\tilde{I}_{i`})$$

"**Background Modeling for Generative Image Models**"
Sandro Schönborn, Bernhard Egger, Andreas Forster, and Thomas Vetter Computer Vision and Image Understanding, Vol 113, 2015.

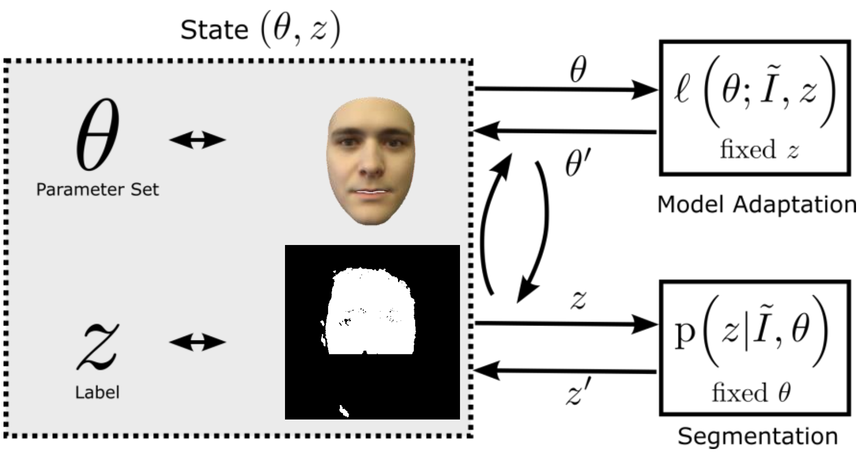
UNIVERSITÄT BASEL
> DEPARTMENT OF MATHEMATICS AND COMPUTER SCIENCE GRAVIS2018 ¦ BASEL

Occlusion-aware Model



$$l(\theta;\tilde{I},z) = \prod_i l_{face}(\theta;\tilde{I}_i)^z \cdot l_{non-face}(\theta;\tilde{I}_i)^{1-z}$$

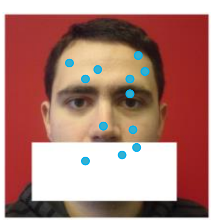
Inference

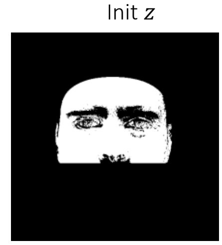


UNIVERSITÄT BASEL
> DEPARTMENT OF MATHEMATICS AND COMPUTER SCIENCE
GRAVIS2018 | BASEL

Initialisation: Robust Illumination Estimation

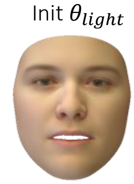
Init z





Init θ_{camera}

Init θ_{light}

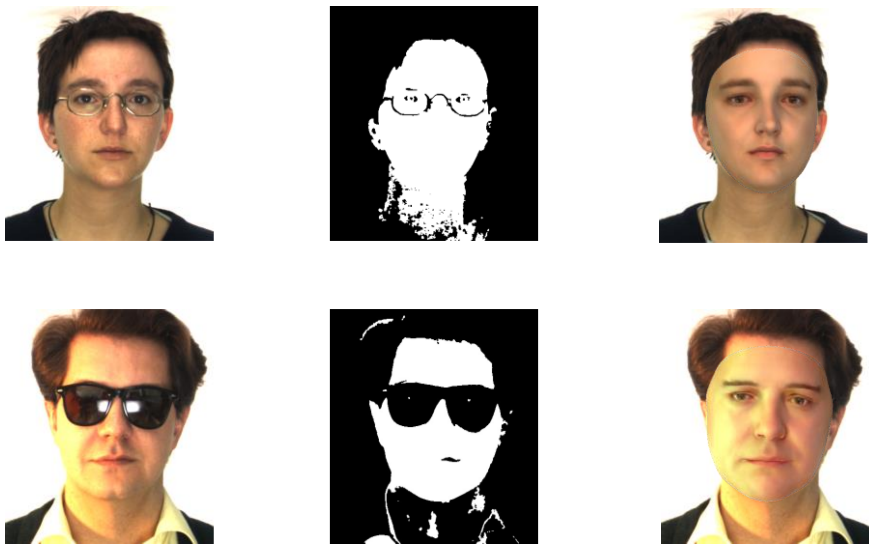


UNIVERSITÄT BASEL

> DEPARTMENT OF MATHEMATICS AND COMPUTER SCIENCE | GRAVIS2018 ¦ BASEL

Results: Qualitative

Source: AR Face Database



UNIVERSITÄT BASEL

> DEPARTMENT OF MATHEMATICS AND COMPUTER SCIENCE | GRAVIS2018 ¦ BASEL

Results: Qualitative

Source: AFLW Database

57

Results: Applications

Source: LFW Database