Occlusion-aware 3D Morphable Models and an Illumination Prior for Face Image Analysis

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

Abstract Faces in natural images are often occluded by a variety of objects. We propose a fully automated, probabilistic and occlusion-aware 3D Morphable Face Model adaptation framework following an Analysis-by-Synthesis setup. The key idea is to segment the image into regions explained by separate models. Our framework includes a 3D Morphable Face Model, a prototype-based beard model and a simple model for occlusions and background regions. The segmentation and all the model parameters have to be inferred from the single target image. Face model adaptation and segmentation are solved jointly using an expectation–maximization-like procedure. During the E-step, we update the segmentation and in the M-step the face model parameters are updated. For face model adaptation we apply a stochastic sampling strategy based on the Metropolis-Hastings algorithm. For segmentation, we apply Loopy Belief Propagation for inference in a Markov random field. Illumination estimation is critical for occlusion handling. Our combined segmentation and model adaptation needs a proper initialization of the illumination parameters. We propose a RANSAC-based robust illumination estimation technique. By applying this method to a large face image database we obtain a first empirical distribution of real-world illumination conditions. The obtained empirical distribution is made publicly available and can be used as prior in probabilistic frameworks, for regularization or to synthesize data for deep learning methods.

1 Introduction

Face image analysis is a major field in computer vision. We focus on 3D reconstruction of a face given a single still image under occlusions. Since the problem of reconstructing a 3D shape from a single 2D image is inherently ill-posed, a strong object prior is required. Our approach builds on the Analysis-by-Synthesis strategy. We extend the approach by an integrated segmentation into face, beard and non-face to handle occlusions.

We work with a 3D Morphable Face Model (3DMM) as originally proposed by Blanz and Vetter (1999) representing the face by shape and color parameters. With illumination and rendering parameters we can synthesize facial images. We adapt all parameters to render an image which is as similar as possible to the target. This process is called *fitting*.

Most "in the wild" face images depict occluders like glasses, hands, facial hair, microphones and various other things (see Köstinger et al (2011)). Such occlusions are a challenge in an Analysis-by-Synthesis setting - the fitting procedure of 3DMMs is misled by them. Occlusions make the face model either drift away, or the shape and color parameters are distorted because the face model adapts to those occlusions. Standard methods like e.g. robust error measures are not sufficient for generative face image analysis since they tend to exclude important details like the eyes, eyebrows or mouth region which are more difficult to explain by the face model (see Figure 2). At the same time, it is impossible to detect occlusions without strong prior knowledge. The challenge of occlusion-aware model adaptation is to explain as much as possible by the face model and only

Department of Mathematics and Computer Science
University of Basel
Basel Switzerland
E-mail: bernhard.egger@unibas.ch

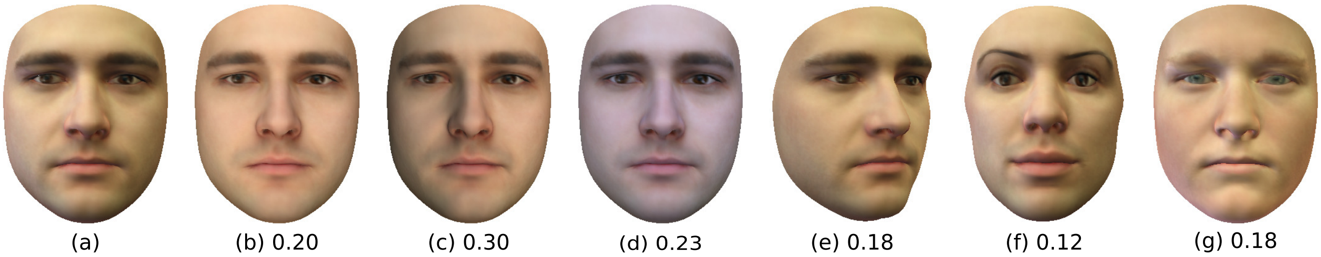

Fig. 1: Illumination dominates facial appearance. We indicate the RMS-distance in color space of different renderings against a target rendering (a). We rendered the same target under new illumination conditions (b-d), compared to other changes (e-g). We present a frontal illumination (b), an illumination from the side (c) and a real-world illumination (d). For comparison, we rendered the original face (a) under the original illumination conditions with strong changes in pose (e), shape (f) and texture (g). All those changes (e-g) are influencing the distance to the target image less than changes in illumination (b-d). The shown RMS distances caused by illumination are on average 50% higher than those caused by changing other parameters.

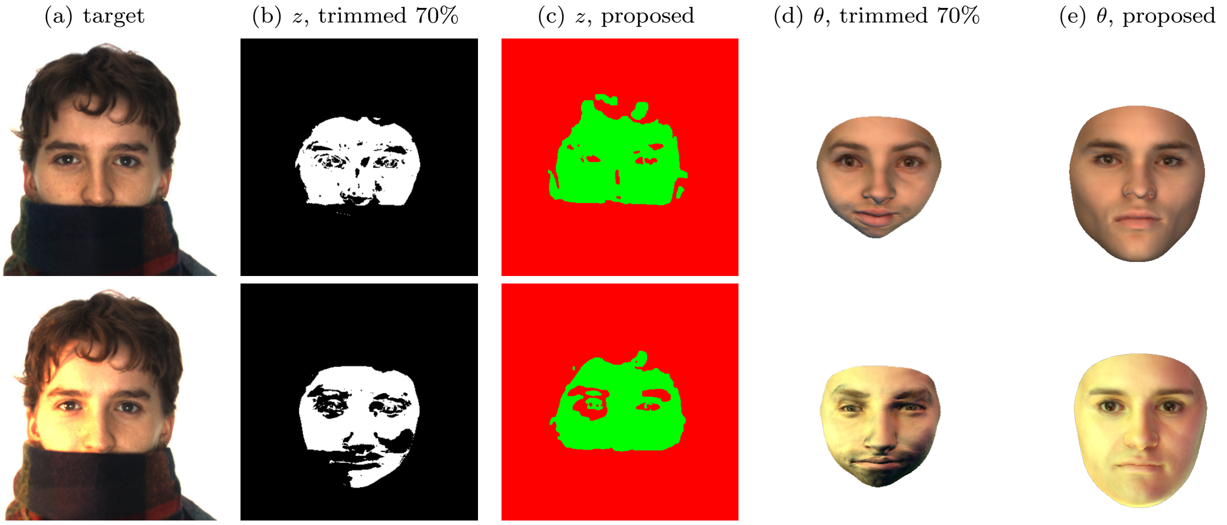

Fig. 2: The effect of illumination on a standard robust technique. (a) depicts the target image from the AR face database by Martinez and Benavente (1998) under homogeneous and inhomogeneous illumination. (b) and (d) are showing the segmentation and fitting result using a trimmed evaluator as described in Egger et al (2016) (considering only the best matching 70% of the pixels). (c) and (e) depicts the segmentation and fitting result of the proposed method. Whilst the segmentation using the trimmed version succeeds only with homogeneous frontal illumination and fails with illumination from the side, our approach is illumination-invariant and succeeds in both cases for segmentation and fitting.

exclude non-face pixels. At the same time, it should not tend to include occlusions or strong outliers.

In the proposed framework we distinguish between occlusions that are specific for faces like beards or glasses and occlusions that are not face specific. Face-specific occlusions are suitable to be modeled explicitly, whereas other objects like microphones, food or tennis rackets are captured by a general occlusion model. In this way, we omit to model all objects that happen to occur near faces explicitly. Beards cannot be handled as general occlusion since the illumination model can explain them by a darker illumination. The beard region tends to mislead illumination estimation and the model adaptation. Another challenge with beards is that they share their appearance with parts included in the model like the eyebrows. Compared to other occlusions, beards are a part of the face and therefore even occluding the face in a cooperative setting. In this paper, we model beards explicitly and handle other occlusions with a general occlusion model. The beard prior is composed of a prototype based shape model and an appearance model derived from the beard region in the target image.

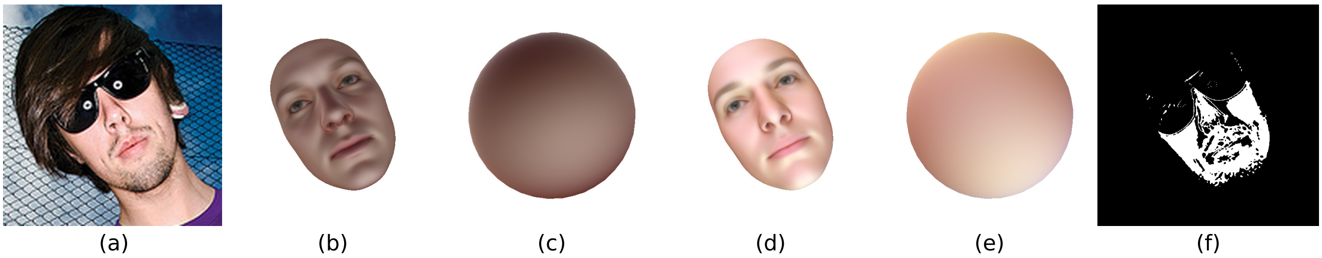

Fig. 3: The target image (a) contains strong occlusions through sunglasses and facial hair. Non-robust illumination estimation techniques are strongly misled by those occlusions. We present the non-robustly estimated illumination (Schönborn et al (2017)) rendered on the mean face (b) and on a sphere with average face albedo (c). In (d) and (e) we present the result obtained with our robust illumination estimation technique. The white pixels in (f) are pixels selected for illumination estimation by our robust approach. The target image is from the AFLW database (Köstinger et al (2011)).

Our approach is to segment the image into face, beard and non-face regions. The beard model is coupled to the face model via its location. Non-face regions can arise from occlusions or outliers in the face or background region. We formulate an extended Morphable Face Model that deals explicitly with beards and other occlusions. This enables us to adapt the face model only to regions labeled as face. The beard model is defined on the face surface in 3D and therefore its location in a 2D image is strongly coupled to the face model. This coupling allows the beard model to guide the face model relative to its location. During inference, we also update the segmentation into face, beard and non-face regions. Segmentation and parameter adaptation cannot be performed separately. We require a set of face model parameters as prior for segmentation and a given segmentation for parameter adaptation. We, therefore, use an expectation–maximization-like (EM) algorithm to perform segmentation and parameter adaptation in alternation.

Previous work on handling occlusions with 3DMMs was evaluated on controlled, homogeneous illumination only. However, complex illuminations are omnipresent in real photographs and illumination determines facial appearance, see Figure 1. We analyze “in the wild” facial images. They contain complex illumination settings and their occlusions cannot be handled with previous approaches. We show the corrupting effect of illumination using a standard robust technique for face model adaptation in Figure 2. We also present the effect of occlusions on non-robust illumination estimation techniques in Figure 3. Therefore, we propose a robust illumination estimation. We incorporate the face shape prior in a RANSAC-like algorithm for robust illumination estimation. The results of this illumination estimation are illumination parameters and an initial segmentation into face and non-face regions. Based on this robust illumination estimation we initialize our face model parameters and the segmentation labels.

The Annotated Facial Landmarks in the Wild database (AFLW) (Köstinger et al (2011)) provides “in the wild” photographs under diverse illumination settings. We estimate the illumination conditions on this database to obtain an unprecedented prior on natural illumination conditions. The prior spans the space of real-world illumination conditions. The obtained data are made publicly available. This prior can be integrated into probabilistic image analysis frameworks like Schönborn et al (2017); Kulkarni et al (2015).

Furthermore, the resulting illumination prior is helpful for discriminative methods which aim to be robust against illumination. Training data can be augmented (Jourabloo and Liu (2016); Zhu et al (2016)) or synthesized (Richardson et al (2016); Kulkarni et al (2015)) using the proposed illumination prior. This is especially helpful for data-greedy methods like deep learning. Those methods are already including a 3D Morphable Model as prior for face shape and texture. Currently, no illumination prior learned on real-world data is available. The proposed illumination prior is an ideal companion of the 3D Morphable Model, allows the synthesis of more realistic images and is first used in a publicly available data generator[1] by Kortylewski et al (2017).

An overview of our full occlusion-aware face image analysis framework is depicted in Figure 4. The paper focuses on the three blocks of robust illumination estimation which can be applied independently or be used as initialization, the full combined segmentation and face model adaptation framework and an illumination prior derived from real-world face images.

[1] `https://github.com/unibas-gravis/parametric-face-image-generator`

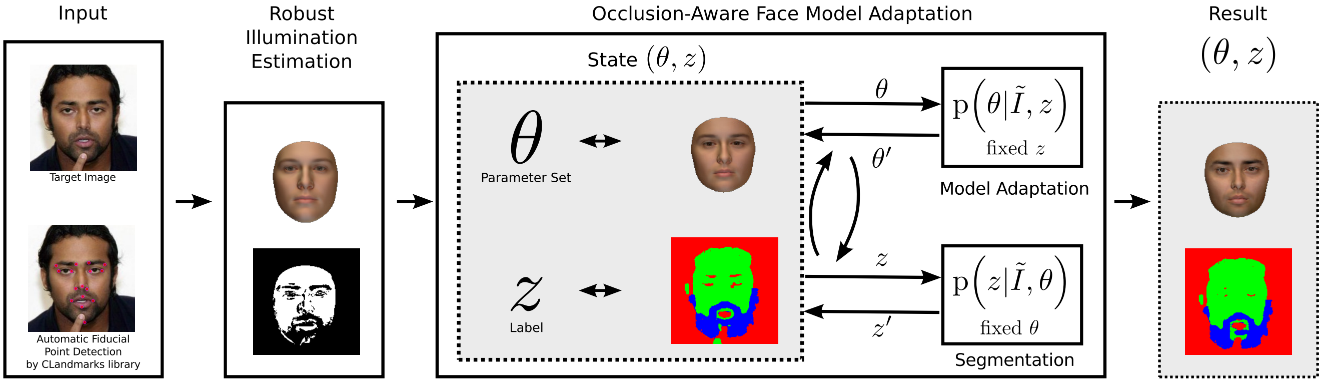
Input
Target Image
Automatic Fiducial Point Detection by CLandmarks library
Robust Illumination Estimation
Occlusion-Aware Face Model Adaptation
State (θ, z)
θ
Parameter Set
z
Label
θ
θ'
z
z'
$p\left(\theta|\tilde{I}, z\right)$
fixed z
Model Adaptation
$p\left(z|\tilde{I}, \theta\right)$
fixed θ
Segmentation
Result
(θ, z)

Fig. 4: Algorithm overview: As input, we need a target image and fiducial points. We use the external Clandmark Library for automated fiducial point detection from still images (Uřičář et al (2015)). We start with an initial face model fit of our average face with a pose estimation. Then we perform a RANSAC-like robust illumination estimation for initialization of the segmentation label z and the illumination setting (for more details see Figure 9). Then our face model and the segmentation are simultaneously adapted to the target image $\tilde{I}$. The result is a set of face model parameters θ and a segmentation into face and non-face regions. The presented target image is from the LFW face database (Huang et al (2007)).

1.1 Related Work

3DMMs are widely applied for 3D reconstruction in an Analysis-by-Synthesis setting from single 2D images (Blanz and Vetter (1999); Romdhani and Vetter (2003); Aldrian and Smith (2013); Schönborn et al (2013); Zhu et al (2015); Huber et al (2015)). Recently there are methods based on convolutional neural networks arising for the task of 3D reconstruction using 3D Morphable Face Models. Besides the work of Tewari et al (2017) they are only reconstructing the shape and not estimating facial color or the illumination in the scene. The work of Tewari et al. is a model adaptation framework which can be trained end-to-end. It is not robust to occlusions and states this as a hard challenge. Although occlusions are omnipresent in face images, most research using 3DMMs relies on occlusion-free data. There exist only a few approaches for fitting a 3DMM under occlusion. Standard robust error measures are not sufficient for generative face image analysis. Areas like the mouth or eye regions tend to be excluded from the fitting because of their strong variability in appearance (Romdhani and Vetter (2003); De Smet et al (2006)), and robust error measures like applied in Pierrard and Vetter (2007) are highly sensitive to illumination. Therefore, we explicitly cover as many pixels as possible by the face model.

De Smet et al (2006) learned an appearance distribution of the observed occlusion per image. This approach focuses on large-area occlusions like sunglasses and scarves. However, it is sensitive to appearance changes due to illumination and cannot handle thin occlusions. The work relies on manually labeled fiducial points and is optimized to the database.

Morel-Forster (2017) predicts facial-occlusions by hair using a random forest. The detected occlusion are integrated in the image likelihood but stay constant during model adaptation.

The recent work of Saito et al. (Saito et al (2016)) explicitly segments the target image into face and non-face regions using convolutional neural networks. Compared to our method it relies on additional training data for this segmentation. The model adaptation does not include color or illumination estimation and is performed in a tracking setting incorporating multiple frames.

The work of Huang et al (2004) combines deformable models with Markov random fields (MRF) for segmentation of digits. The beard prior in our work is integrated in a similar way as they incorporate a prior from deformable models. The method of Wang et al (2007) also works with a spherical harmonics illumination model and a 3D morphable model. The face is also segmented into face and non-face regions to handle occlusions using MRFs. They work on grayscale images in a highly controlled setting and focus on relighting of frontal face images. They do not show any segmentation results which we can compare to. The idea of having different models competing to explain different regions of the image is related to image parsing framework proposed by Tu et al (2005) and our implementation is unique for face image analysis.

Maninchedda et al (2016) proposed joint reconstruction and segmentation working with 3D input data. The semantic segmentation is also used to improve the quality of face reconstruction. The general challenges are related but they used depth information, which is not available in our setting. Depth information strongly helps when searching for occlusions, beards or glasses.

Previous work on occlusion-aware 3DMM adaptation focussed on databases with artificial and homogeneous, frontal illumination settings. We present a model which can handle occlusion during 3DMM adaptation on illumination conditions arising in “in the wild” databases.

Yildirim et al (2017) presented a generative model including occlusions by various objects. 3D occlusions are included in the training data. During inference, the input image is decomposed into face and occluding object and occlusions are excluded for face model adaptation.

Similar models have recently been proposed besides applications of the 3DMM: In the medical imaging for atlas-based segmentation of leukoaraiosis and strokes from MRI brain images (Dalca et al (2014)) and for model-based forensic shoe-print recognition from highly cluttered and occluded images (Kortylewski (2017)).

Robust illumination estimation or inverse lighting is an important cornerstone of our approach. Inverse lighting (Marschner and Greenberg (1997)) is an inverse rendering technique trying to reconstruct the illumination condition. Inverse rendering is applied for scenes (Barron and Malik (2015)) or specific objects. For faces there are 3D Morphable Models by Blanz and Vetter (1999) as the most prominent technique used in inverse rendering settings. The recent work of Shahlaei and Blanz (2015) focuses on illumination estimation and provides a detailed overview of face-specific inverse lighting techniques. The main focus of the presented methods is face model adaptation in an Analysis-by-Synthesis setting. Those methods are limited either to near-ambient illumination conditions (De Smet et al (2006); Pierrard and Vetter (2007)) or cannot handle occlusions (Romdhani and Vetter (2003); Aldrian and Smith (2013); Schönborn et al (2017)).

Our robust illumination estimation technique handles both, occlusions and complex illumination conditions by approximating the environment map using a spherical harmonics illumination model. Few methods incorporate prior knowledge of illumination conditions. The most sophisticated priors are multivariate normal distributions learned on spherical harmonics parameters estimated from data as proposed in Schönborn et al (2017) and Barron and Malik (2015). Those priors are less general and not available to the research community. Our robust estimation method enables us to estimate an illumination prior from available face databases. This illumination prior fills a gap for generative models.

This work extends our previous work (Egger et al (2016)). We additionally integrate an explicit beard model and a prior during segmentation. The Chan-Vese segmentation (Chan and Vese (2001)) was therefore exchanged by a Markov Random field segmentation since it features multi-class segmentation and can easily integrate a prior on the segmentation label. The explicit modeling of beards improves the quality of fits compared to when beards are handled as occlusion. The robust illumination estimation technique was proposed in our previous work but was not explained in detail, we added a complete description of the algorithm and an evaluation of its robustness. As an application of this work, we obtain and publish a first illumination prior available to the research community.

2 Methods

In this section, we present a method for combining occlusion segmentation and 3DMM adaptation into an occlusion-aware face analysis framework (Figure 4). Our segmentation distinguishes between face, beard, and non-face. The initialization of the segmentation is estimated by a RANSAC strategy for robust illumination estimation.

Our approach is based on five main ideas: First, we exclude occlusions during the face model adaptation. The face model should be adapted only to pixels belonging to the face. Second, we model beards explicitly. The segmentation of the beard on a face can guide face model adaptation. Third, we semantically segment the target image into occlusions, beards and the face. We pose segmentation as a Markov random field (MRF) with a beard prior. Fourth, we robustly estimate illumination for initialization. Illumination is dominating facial appearance and has to be estimated to find occlusions. Fifth, we perform face model adaptation and segmentation at the same time using an EM-like procedure. The face model adaptation assumes a given segmentation and vice-versa.

2.1 Image Model

The 3D Morphable Face Model was first described by Blanz and Vetter (1999). Principal Component Analysis (PCA) is applied to build a statistical model of face shape and color. Faces are synthesized by rendering them with an illumination and camera model. We work with the publicly available Basel Face Model (BFM) presented by Paysan et al (2009). The model was interpreted in a Bayesian framework using Probabilistic PCA by Schönborn et al (2017).

The aim of face model adaptation (fitting) is to synthesize a face image which is as similar to the target image as possible. A likelihood model is used to rate face model parameters θ given a target image. The parameter set θ consists of the color transform, camera (pose), illumination and statistical parameters describing the face. The likelihood model is a product of the pixels i of the target image $\tilde{I}$. In the formulation of Schönborn et al (2017), pixels belong to the face model ($\mathcal{F}$) or the background model ($\mathcal{B}$). The foreground and background likelihoods (ℓ_{face}, b) compete to explain pixels in the image. The full likelihood model covering all pixels i in the image is

$$\ell\left(\theta;\tilde{I}\right) = \prod_{i\in\mathcal{F}} \ell_{\text{face}}\left(\theta;\tilde{I}_i\right) \prod_{i'\in\mathcal{B}} b\left(\tilde{I}_{i'}\right). \tag{1}$$

The foreground F is defined solely by the position of the face model (see Figure 5) and therefore this formulation cannot handle occlusions.

2.2 Occlusion Extension of the Image Model

We extend (1) to handle occlusion. We distinguish between beards and other sources of occlusions. Therefore, we introduce a random variable z, indicating the class k a pixel belongs to. The standard likelihood model (1) is therefore extended to incorporate different classes:

$$\ell\left(\theta;\tilde{I},z\right) = \prod_{i}\prod_{k} \ell_k\left(\theta;\tilde{I}_i\right)^{z_{ik}} \tag{2}$$

with $\sum_k z_{ik} = 1\ \forall i$ and $z_{ik} \in \{0,1\}$.

The likelihood model is open for various models for different parts of the image. In this work we use three classes k, namely face (z_{face}), beard (z_{beard}) and non-face ($z_{\text{non-face}}$). In Figure 5 we present all different labels and regions.

The main difference to the formulation by Schönborn et al (2017) is that the face model does not have to fit all pixels in the face region. Pixels in the image are evaluated by different likelihoods ℓ_k for the respective class models k.

For face model adaptation we chose the strongest label z for every pixel $\max_k z_{ik}$. The generative face model is adapted to pixels with the label z_{face} only, according to (2). Beard and other non-face pixels are handled by separate likelihoods during face model adaptation. Non-face pixels are only characterized by a low face and beard likelihood. Thus, they can be outliers, occlusions or background pixels. A graphical model of the full occlusion-aware Morphable Model is depicted in Figure 6.

2.3 Segmentation

To estimate the label z for a given parameter set θ we used an extension of the classic Markov random field segmentation technique including a beard prior similar as in Huang et al (2004), see Figure 8. The beard prior c is a prior on the labels z: $P(z|\theta,c)$.

The beard prior is a simple prior based on m prototype shapes $l \in \{1..m\}$. We use an explicit beard prior, which is defined on the face model (see Figure 7). In the image, it is depending on the current camera and shape model parameters contained θ. For the non-face and face label, we use a uniform prior.

The MRF is formulated as follows:

$$P(z|\tilde{I},\theta) \propto$$
$$\prod_{i}\prod_{k} \ell_k\left(\theta;\tilde{I}_i\right)^{z_{ik}} P(z_{ik}|\theta)P(c) \prod_{j\in n(i)} P(z_{ik}, z_{jk}). \tag{3}$$

The data-term is built from the likelihoods for all classes k and overall pixels i and combined with the beard prior. The smoothness assumption enforces spatial contiguity of all pixels j which are neighbors $n(i)$ of i.

The prior for beards is a prior over the segmentation label z at a certain pixel position i. The label z is depending on the beard prior $P(c)$. The prior is based on prototype shapes defined on the face surface. The segmentation prior c is derived from the current set of face model parameters θ since the pose and shape of the face influence its position and shape in the image. We derived the prototype from manual beard segmentations labeled on the Multi-PIE database (Gross et al (2010)). We used k-means++ clustering technique as proposed in Arthur and Vassilvitskii (2007) to derive a small set of prototypes. The resulting prototypes are shown in Figure 7. We manually added a prototype for no-beard and another one to handle occlusion of the complete beard region. Those additional prototypes allow us to consider all possible labels in the beard region of the face. All prototypes are defined on the face surface and their position in the image is depending on the current face model parameters θ.

2.4 Likelihood Models

Depending on the class label z we apply different likelihood models during model adaptation for each pixel. The same likelihoods are used for segmentation.

2.4.1 Face Likelihood

The likelihood of pixels to be explained by the face model is the following:

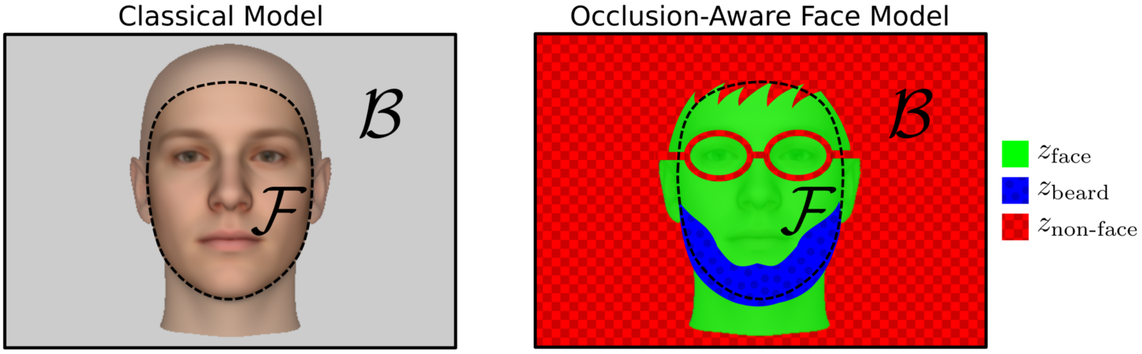

Fig. 5: On the left, the regions used by the likelihood model by Schönborn et al (2013). Each pixel belongs to the face model region $\mathcal{F}$ or the background model region $\mathcal{B}$. Assignment to foreground or background is based on the face model visibility only. In the proposed framework we have the same labels $\mathcal{F}$ and $\mathcal{B}$ but additional segmentation variables z to integrate occlusions as shown on the right. We assign a label z indicating if the pixel belongs to face, beard or non-face. Occlusions in the face model region $\mathcal{F}$ (in this case glasses) can hereby be excluded from the face model adaptation. Beards are handled explicitly and labeled separately.

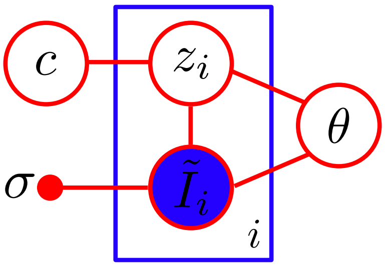

Fig. 6: Graphical model of our occlusion-aware Morphable Model. We observe the target image $\tilde{I}$ under Gaussian noise σ. To explain different pixels i by different models we introduce a class segmentation label z for each pixel. The segmentation uses a beard prior c which is explained in more detail in Section 2.3. The pixels labeled as face are explained by the face model parameters θ.

$$\ell_{\text{face}}\left(\theta;\tilde{I}_i\right) = \begin{cases} \frac{1}{N}\exp\left(-\frac{1}{2\sigma^2}\left\|\tilde{I}_i - I_i(\theta)\right\|^2\right) & \text{if } i \in \mathcal{F} \\ \frac{1}{\delta}h_f(\tilde{I}_i,\theta) & \text{if } i \in \mathcal{B}. \end{cases} \tag{4}$$

Pixels are evaluated by the face model if they are labeled as face (z_{face}) and are located in face region $\mathcal{F}$. The rendering function I generates a face for given parameters θ. This synthesised image $I_i(\theta)$ is compared to the observed image $\tilde{I}$. The likelihood model for pixels covered by the face model is assuming per-pixel Gaussian noise in the face region. The likelihood ℓ_{face} is defined over the whole image and therefore also in the non-face region $\mathcal{B}$. For those pixels that are not covered by the generative face model, we use a simple color model to compute the likelihood of the full image. We use a color histogram h_f with δ bins estimated on all pixels in $\mathcal{F}$ labelled as face (z_{face}).

2.4.2 Beard Likelihood

The beard likelihood is a simple histogram color model. The histogram h_b with δ bins, is estimated on the current belief on z_{beard}:

$$\ell_{\text{beard}}\left(\theta;\tilde{I}_i\right) = \frac{1}{\delta}h_b(\tilde{I}_i,\theta). \tag{5}$$

In Egger (2017) we also proposed a discriminative approach based on hair detection as beard likelihood. This likelihood can be extended to more elaborate appearance and shape models as proposed in Le et al (2015) or Nguyen et al (2008).

2.4.3 Occlusion and Background Likelihood

The likelihood of the non-face model to describe occluding and background pixels is the following:

$$\ell_{\text{non-face}}\left(\theta;\tilde{I}_i\right) = b\left(\tilde{I}_i\right) = \frac{1}{\delta}h_{\tilde{I}}(\tilde{I}_i). \tag{6}$$

This likelihood allows us to integrate simple color models based on the observed image $\tilde{I}$. We use a simple color histogram estimated on the whole image $\tilde{I}$ to

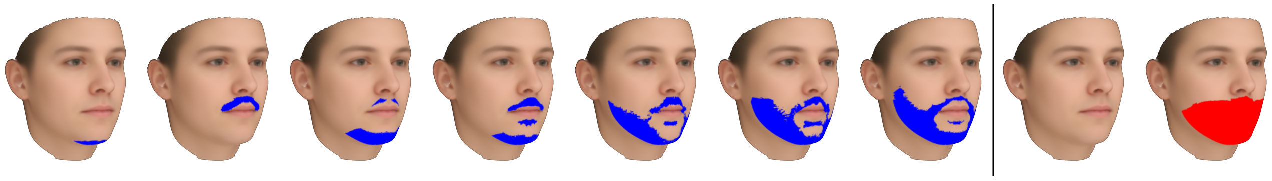

Fig. 7: The seven beard prototypes derived from k-means++ clustering on manual beard segmentations on the Multi-PIE database (blue labels). We manually added a prototype for no-beard and one to handle occlusions over the complete beard region (right). The prototypes are defined on the 3D face model and can be rendered to the image according to the face model parameters θ.

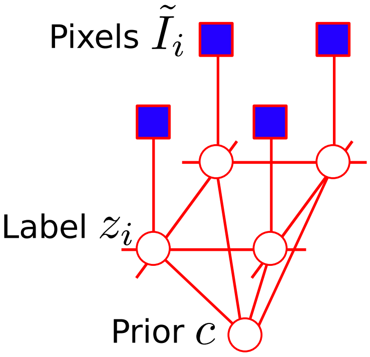

Fig. 8: Graphical model of our Markov Random field with beard prior c. We are interested in the segmentation labels z_i and observe the image pixels $\tilde{I}_i$. Every label z_i at position i is connected to the prior c and its neighbours.

estimate the background likelihood as described by Egger et al (2014).

2.5 Inference

The full model consists of the likelihoods for face model adaptation shown in (2) and segmentation from (3). Those equations depend on each other. To estimate the segmentation label z we assume a given set of face model parameters θ. And vice versa we assume a given segmentation label z to adapt the face model parameters θ. Both are not known in advance and are adapted during the inference process to get a MAP-estimate of the face model parameters and the segmentation. We use an EM-like algorithm (Dempster et al (1977)) for alternating inference of the full model. In the expectation step, we update the label z of the segmentation. In the maximisation step, we adapt the face model parameters θ. An overview of those alternating steps is illustrated in Figure 4.

Face model adaptation is performed by a Markov Chain Monte Carlo strategy (Schönborn et al (2017)) with our extended likelihood from (2). MRF segmentation is performed using Loopy Belief Propagation with a sum-product algorithm as proposed by Murphy et al (1999).

The histogram-based appearance model of beards is adapted during segmentation and fitting. During the segmentation, the appearance is updated respecting the current belief on z_i. During the fitting, the beard appearance is also updated due to the spatial coupling with the face model. When the shape or camera parameters of the face model change, the beard model has to be updated.

During segmentation, we assume fixed parameters θ and during fitting, we assume given labels $\max_k z_{ik}$. Since the fixed values are only an approximation during the optimisation process, we account for those uncertainties. These misalignments are especially important in regions like the eye, nose, and mouth. These regions are often mislabelled as occlusion due to their high variability in appearance when using other robust error measures. In the inference process, those regions are automatically incorporated gradually by adapting the face and non-face likelihood to incorporate this uncertainty. To account for the uncertainty of the face model parameters θ during segmentation, we adapt the face model likelihood for segmentation (compare to (4)):

$$\ell'_{\text{face}}(\theta; \tilde{I}_i) = \frac{1}{N} \exp\left(-\frac{1}{2\sigma^2} \min_{j \in n(i)} \left\| \tilde{I}_i - I_{i,j}(\theta) \right\|^2 \right). \tag{7}$$

The small misalignment of the current state of the fit is taken into account by the neighbouring pixels j in the target image. In our case we take the minimum over a patch of the 9×9 neighbouring pixels direction (interpupillary distance is ˜120 pixels).

To account for the uncertainty of the segmentation label z for face model adaptation, we adapt the likelihood of the non-face during face model adaptation (compare to (6)). Pixels which are masked as non-face can be explained by the face model if it can do better:

$$\ell'_{\text{non-face}}\left(\theta,\tilde{I}_i\right)=\max\left(\ell_{\text{face}}\left(\theta,\tilde{I}_i\right),b\left(\tilde{I}_i\right)\right)\text{if } i\in\mathcal{F}. \tag{8}$$

Both modifications more likely label pixels as face and this leads to consider them during face model adaptation.

2.6 Initialisation and Robust Illumination Estimation

In the early steps of face model adaptation under occlusion, we need an initial label z. Occlusions are however hard to determine in the beginning of the face model adaptation due to the strong influence of illumination on facial appearance (see Figure 1 and Figure 2).

Algorithm 1: Robust Illumination Estimation.

```
Input: Target image $\tilde{I}$, surface normals $\mathbf{n}$, albedo $a^c$,
       iterations $m$, number of points for illumination
       estimation $n$, threshold $t$, minimal size $k$ of
       consensus set, rendering function $I$
Output: Illumination parameters $L_{lm}$, consensus set $c$

$c = \varnothing$
for $m$ iterations do
    1. Draw $n$ random surface points $p$
    2. Estimate $L_{lm}$ on $p, \mathbf{n}$, $a^c$ and $\tilde{I}$ (Eq. 9)
    3. Compare $I(L_{lm})$ to $\tilde{I}$ (Eq. 4). Pixels consistent
       with $L_{lm}$ ($\ell_{\mathrm{L}}(L_{lm};\tilde{I}_i) > t$), build $c$.
    4. if $|c| > k$ then
        Estimate illumination on $c$
        Save $c$ if $L_{lm}$ is better than previous best
```

The main requirement for our illumination estimation method is robustness against occlusions. The idea is to find the illumination setting which most consistently explains the observed face in the image. With a set of points with known albedo and shape, we estimate the illumination condition.

Occlusions and outliers render this task ill-posed and mislead non-robust illumination techniques. The selection of the points used for the estimation shall not contain outliers or occlusions. We use an adapted random sample consensus algorithm (RANSAC) (Fischler and Bolles (1981)) for robust model fitting. We synthesize illumination conditions estimated on different point sets and find the illumination parameters most consistent with the target image. The following steps of our procedure are visualized in Figure 9 and written in Pseudo-Code in Algorithm 1.

The idea of the RANSAC algorithm adapted to our task is, to randomly select a set of points on the face surface which are visible in the target image $\tilde{I}$ and estimate the illumination parameters L_{lm} from the appearance of those points (step 1 and 2 of Algorithm 1). The estimated illumination is then evaluated on all available points (step 3). The full set of points consistent with this estimation is called the consensus set. Consistency is measured by counting the pixels of the target image which are explained well by the current illumination setting. If the consensus set contains more than 40% of the visible surface points the illumination is re-estimated on 1000 random points from the full consensus set for a better approximation. If this illumination estimation is better than the last best estimation according to Equation (4), it is set as the currently best estimation.

The sampling on the face surface (step 1 of Algorithm 1) considers domain knowledge. There are some occlusions which often occur on faces like facial hair or glasses. Therefore we include a region prior to sample more efficiently in suitable regions. Details on how to obtain this prior are in Section 2.6.2.

Most "in the wild" face images contain compression artifacts (e.g. from the jpeg file format). To reduce those artifacts and noise we blur the image in a first step with a Gaussian kernel ($r = 4$ pixels with an image resolution of 512×512 pixels).

2.6.1 Illumination Model

The proposed algorithm is not limited to a specific illumination model. The main requirement is, that it should be possible to estimate the illumination from a few points with given shape, albedo, and appearance. The spherical harmonics illumination model has two main advantages. First, it is able to render natural illumination conditions by approximating the environment map. Second, the illumination estimation from a set of points is solved in a system of linear equations.

Spherical harmonics allow an efficient representation of an environment map with a small set of parameters L_{lm} (Ramamoorthi and Hanrahan (2001); Basri and Jacobs (2003)). The radiance function is parametrized through real spherical harmonics basis functions Y_{lm}. The radiance p_j^c per color channel c and for every point j on the surface is derived from its albedo a_j and surface normal $\mathbf{n}_j$ and the illumination parameters L_{lm}:

$$p_j^c = a_j^c \sum_{l=0}^{2} \sum_{m=-l}^{l} Y_{lm}(\mathbf{n}_j) L_{lm}^c \alpha_l. \tag{9}$$

The expansion of the convolution with the reflectance kernel is given by α_l, for details, refer to Basri and Jacobs (2003). We use Phong shading and interpolate

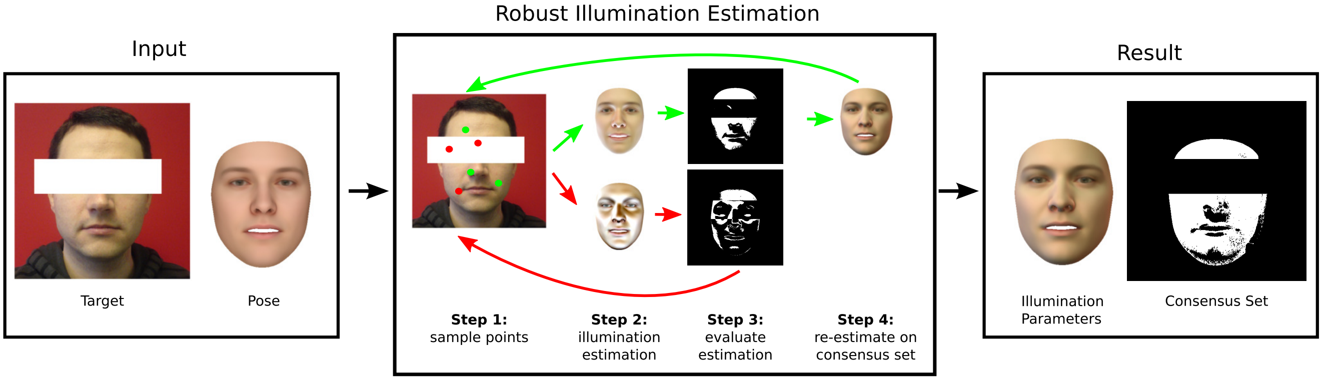

Fig. 9: Robust Illumination Estimation: Our RANSAC-like algorithm (compare Algorithm 1) takes the target image and a pose estimation as input. We added a strong occlusion (white bar) for better visualization. The algorithm samples iteratively points on the face surface (step 1). We estimate the illumination from the appearance of those points in the target image (step 2). The estimation is then evaluated by comparing the model color to the target image and threshold its likelihood (step 3). The illumination is misled by including the occluded regions (red points). Choosing good points (green points) leads to a reasonable estimation. If the consensus set is big enough we re-estimate the illumination on the full consensus set. We repeat the estimation for different point sets and the most consistent one is chosen as a result.

the surface normal for each pixel. As the light model is linear (Equation 9), the illumination expansion coefficients L^c_{lm} are estimated directly by solving a linear system (least squares) with given geometry, albedo and observed radiance as described by Zivanov et al (2013). This system of linear equations is solved during the RANSAC algorithm using a set of vertices j.

2.6.2 Region Prior

Facial regions differ in appearance and elicit strong variations. Whilst some regions like the eyebrows or the beard region vary strongly between different faces, other regions like the cheek stay more constant. Also, common occlusions through glasses or beards strongly influence facial appearance. Regions with low appearance variation are more suitable for illumination estimation than those with stronger variation. We restrict the samples in the first step of the RANSAC algorithm to the most constant regions. The regions with strong variation are excluded for sample generation (step 1 of Algorithm 1) but included in all other steps.

We estimate the texture variation on the Multi-PIE database (Gross et al (2010)) which contains faces with glasses and beards under controlled illumination. We chose the images with frontal pose (camera 051) and with frontal, almost ambient illumination (flash 16) from the first session. We chose this subset to exclude all variation in illumination and pose (which we model explicitly) for our prior. The variation is estimated on all 330 identities. The images are mapped on the face model surface by adapting the Basel Face Model to each image with the approach by Schönborn et al (2017). For the first step of the illumination estimation algorithm, we use the regions of the face where the texture variation is below half of the strongest variation. The variation and the resulting mask are depicted in Figure 10.

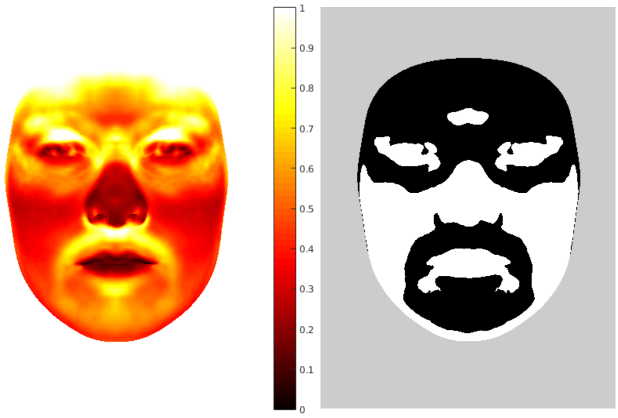

Fig. 10: We derive a region prior from the average variance of appearance overall color channels per vertex on the face surface (on the left), details in Section 2.6.2. Scaling is normalized by the maximal observed variance. On the right, the mask obtained by thresholding at half of the maximal observed variance. We use the white regions to sample points in the first step of our algorithm. Note that especially multi-modal regions which do or do not contain facial hair or glasses are excluded.

2.6.3 Illumination and Segmentation Initialisation

Our robust illumination estimation gives a rough estimate of the illumination and segmentation. However, the obtained mask is underestimating the face region. Especially the eye, eyebrow, and mouth regions are not included in this first estimate. Those regions differ from the skin regions of the face by their higher variance in appearance, they will be gradually incorporated during the full model inference.

The initialization of the beard model is derived from the segmentation obtained by robust illumination estimation. The prior is initialized by the mean of all beard prototypes. The appearance is estimated from the pixels in the prototype region segmented as non-face by the initial segmentation.

3 Experiments and Results

We present results measuring the performance of our initialization and the full model adaptation separately. The robust illumination estimation plays an important role since later steps rely on a proper initialization. For the full model, we separately evaluate the segmentation and quality of the fit.

The software implementation is based on the Statismo (Lüthi et al (2012)) and Scalismo[2] software frameworks. Our source code will be released within the Scalismo-Faces project[3]. We also prepared tutorials with working examples[4].

3.1 Robustness of Illumination Estimation

To evaluate the robust illumination estimation we performed our experiments on faces using a spherical harmonics illumination model, however, the algorithm is open to various objects and illumination models. We use synthetic data to evaluate the sensitivity against occlusion. Our algorithm assumes a given pose, we investigate the errors introduced by this pose estimation. We exclude appearance variations by using a simple prior for shape and texture. We also examine the sensitivity to shape and texture changes. Besides the quantitative evaluation on synthetic data in this Section, the performance of our method is evaluated on real-world face images in a qualitative experiment in Section 3.1.1. And last, we present a novel illumination prior learned on empirical data in Section 3.1.2.

We rely on the mean face of the Basel Face Model (Paysan et al (2009)) as prior for face shape and texture for all experiments. We used $n = 30$ points for the illumination estimation as parameters in step 2. We use $\sigma = 0.043$ estimated on how well the Basel Face Model is able to explain a face image (Schönborn et al (2017)) and threshold the points for the consensus set at $t = 2\sigma$. We estimate the illumination on the full consensus set if the consensus set contains more than $x = 40\%$ of the surface points visible in the rendering. We stop the algorithm after $m = 500$ iterations.

We measure on synthetic data how much occlusion our algorithm is able to handle. We also investigate how robust the algorithm is against pose misalignments and how much our simplified shape and texture prior influences the result. We need ground truth albedo and illumination and therefore generate synthetic data. We use the mean shape and texture of the Basel Face Model as an object and render it under 50 random illumination conditions.

For the first experiment, we add a synthetic random occlusion to this data. The random occlusion is a block with a random color. Note this type of occlusions are worst-case occlusions. The synthetic occlusions are positioned randomly on the face. We randomly generate illumination parameters L_{lm} to synthesize illumination conditions. We sample the parameters according to a uniform distribution between -1 and 1 for all parameters. An example of the synthesized data is depicted in Figure 11. We estimate the illumination condition on these data using our robust illumination estimation technique to measure the robustness. We measured the approximation error by measuring the RMS-distance in color space of the sphere rendered with the estimated illumination condition and the sphere with the ground truth illumination condition as proposed by Barron and Malik (2015).

We cope with 40% of occlusion and reach a constantly good estimation, see Figure 12. Occlusions which surpass 40% and which can partially be explained by illumination, are not properly estimated by our algorithm, see Figure 11. The generated illumination conditions are unnatural, therefore we also evaluated the robustness against occlusion on observed real-world illumination conditions (Section 3.1.1). Every measurement is based on 50 estimations.

Our algorithm relies on a given pose estimation. Pose estimation is a problem which can only be solved approximatively. We, therefore, show how robust our algorithm is against small misalignments in the pose.

[2] Scalismo - A Scalable Image Analysis and Shape Modelling Software Framework available as Open Source under `https://github.com/unibas-gravis/scalismo`

[3] Scalismo-Faces - Famework for shape modeling and model-based image analysis available as Open Source under `https://github.com/unibas-gravis/scalismo-faces`

[4] Tutorials on our Probabilistic Morphable Model framework http://gravis.dmi.unibas.ch/PMM/

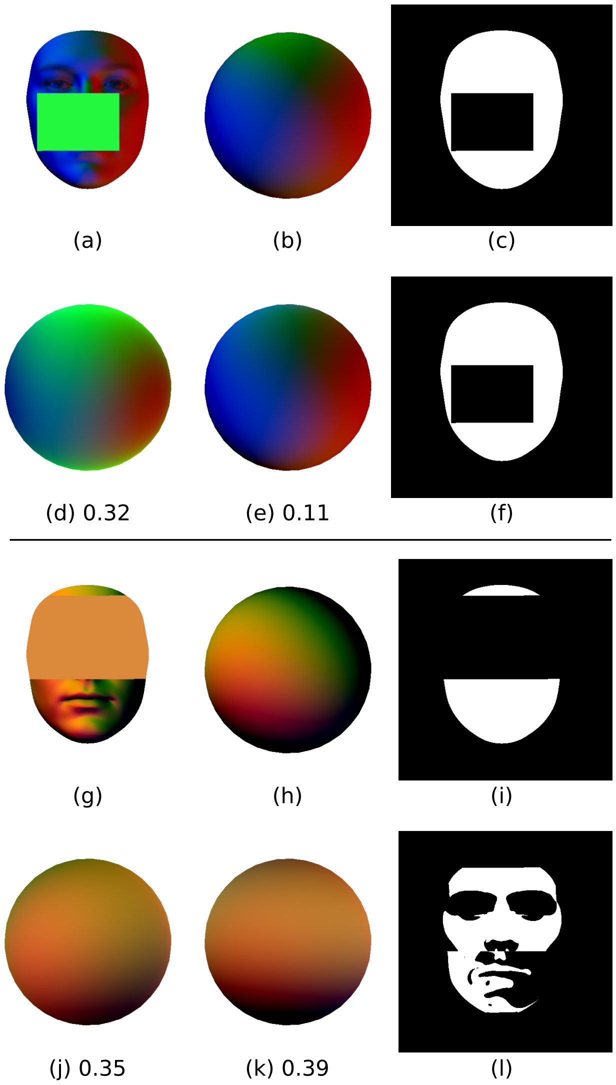

Fig. 11: Two examples of our synthetic data to measure the robustness against occlusions of our approach. In the successful case (a-f), the consensus set is perfect and the occlusion is excluded from the illumination estimation. In the failed case (g-l) the chosen occlusion is similar in color appearance to the observed face appearance. This leads the best consensus set to fail in explaining the correct region of the image. The first example (a-f) is a synthesized target image with 30% of the face occluded. The second example is with 60% occlusion (g-l). The target image is shown in (a, g) The ground truth illumination is rendered on a sphere for comparison (b, h). The ground truth occlusion map is rendered in (c, i). The baseline illumination estimation estimated on 1000 random points is shown in (d, j). Our robust illumination estimation result (e, k), as well as the estimated mask, is shown in (f, l). Together with the visual result, we indicate the measured RMS-distance on the rendered sphere in color space.

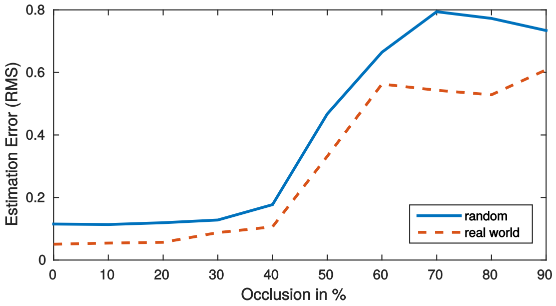

Fig. 12: We measured the illumination estimation error of our approach related to the degree of occlusion. We compare randomly generated illumination conditions with those arising from real world settings. We observe that our algorithm is robust until 40% of occlusion.

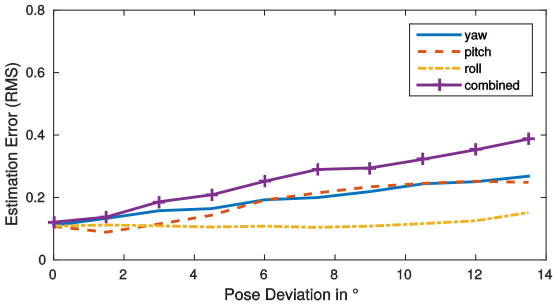

Fig. 13: We measured the illumination estimation error of our approach related to pose estimation errors. Our algorithm handles the pose deviations which arise by wrong pose estimation as input.

We again generate synthetic data with random illumination parameters and manipulate the pose before we estimate the illumination. The results are shown in Figure 13. We present the separate effects of yaw, pitch, and roll as well as combining all three error sources. Small pose estimation errors still lead to good estimations, as expected they grow with stronger pose deviations. We have to expect errors smaller than 10 degrees from pose estimation methods (compare Murphy-Chutorian and Trivedi (2009)).

We use a simple prior for shape and texture for the proposed illumination estimation. In this experiment, we want to measure how changes in shape and texture influence the illumination estimation. We, therefore, manipulate all shape respectively color parameters by a fixed standard deviation. The result is presented in Figure 14. We observe that those variations influence the illumination estimation but do not break it.

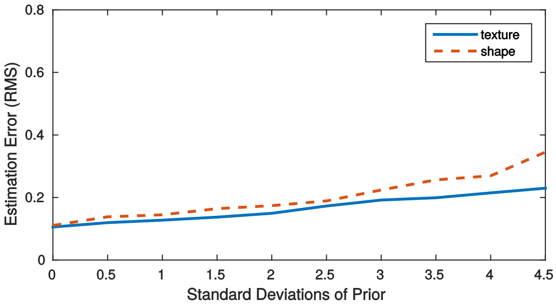

Fig. 14: We measured the illumination estimation error of our approach related to shape and texture changes. Even with a very simple appearance prior we reasonably estimate the illumination condition.

3.1.1 Illumination Estimation in the Wild

We applied our robust illumination estimation method on the Annotated Facial Landmarks in the Wild database (AFLW) (Köstinger et al (2011)) containing 25'993 images with a high variety of pose and illumination. For this experiment and the construction of the illumination prior we used the manual landmarks (all other experiments use CLandmarks) provided within the database for a rough pose estimation following the algorithm proposed by Schönborn et al (2017). The illumination conditions are not limited to lab settings but complex and highly heterogeneous. We observe that small misalignments due to pose estimation still lead to a reasonable illumination estimation. The affected pixels, e.g. in the nose region, are automatically discarded by the algorithm.

Both, the estimated illumination and the occlusion mask arising from the consensus set can be integrated for further image analysis. We present a selection of images under a variety of illumination conditions with and without occlusions in Figure 15. The illumination estimation results demonstrate the robustness of our approach against occlusions like facial hair, glasses, and sunglasses in real-world images.

3.1.2 Illumination Prior

The idea of an illumination prior is to learn a distribution of natural illumination conditions from training data. There are various areas of application for such a statistical prior. It can be directly integrated into generative approaches for image analysis or be used to synthesize training data for discriminative approaches.

For faces, the 3D Morphable Face Model (Blanz and Vetter (1999), Paysan et al (2009)) provides such a prior distribution for shape and color but it does not contain a prior on illumination. Such priors to generate synthetic data recently attracted the attention of the deep learning community for their generative capabilities. 3D Morphable Face Models were already used as an appearance prior for data synthesis (Richardson et al (2016); Kulkarni et al (2015)) or augmentation (Jourabloo and Liu (2016); Zhu et al (2016)). An additional illumination prior is essential to generate realistic images. We, therefore, publish an illumination prior for the diffuse spherical harmonics illumination model (see Section 2.6.1) learned on real-world illumination conditions.

We derive an illumination prior on real-world face images by applying our proposed robust illumination estimation method described in Algorithm 1 to a large face database. We again chose the AFLW database since it contains a high variety of illumination settings. We applied the pose and robust illumination estimation as described in Section 3.1.1. We excluded gray-scale images and faces which do not match our face model prior (strong make-up, dark skin, strong filter effects). We manually excluded the images where the estimation failed and used the remaining 14'348 images as training data. The parameters for illumination estimation where chosen as described in Section 3.1.

By concatenating all spherical harmonic parameters L_{lm}^c of the first 3 bands and color channels we get 27 parameters. As a model, we propose to use a multivariate normal distribution on the spherical harmonics parameters L_{lm}^c and present the first eigenmodes (PCA) in Figure 16. The illumination conditions are normalized relative to the camera position (not to the facial pose). We show some new unseen random illumination conditions in Figure 17.

Together with this paper, we publish the estimated illumination (empirical distribution) of our illumination prior as well as the multivariate normal distribution. The prior allows us to generate unseen random instances from the prior distribution to synthesize illumination settings.

The limitations of this illumination prior and the robust illumination estimation are directly deduced from the used spherical harmonics illumination model. We did not incorporate specular highlights or effects of self-occlusion explicitly to keep the model simple. This simplification does not mislead our approximation since regions which are sensitive to self-occlusion or contain specular highlights are excluded from the illumination estimation during our robust approach (see Figure 15).

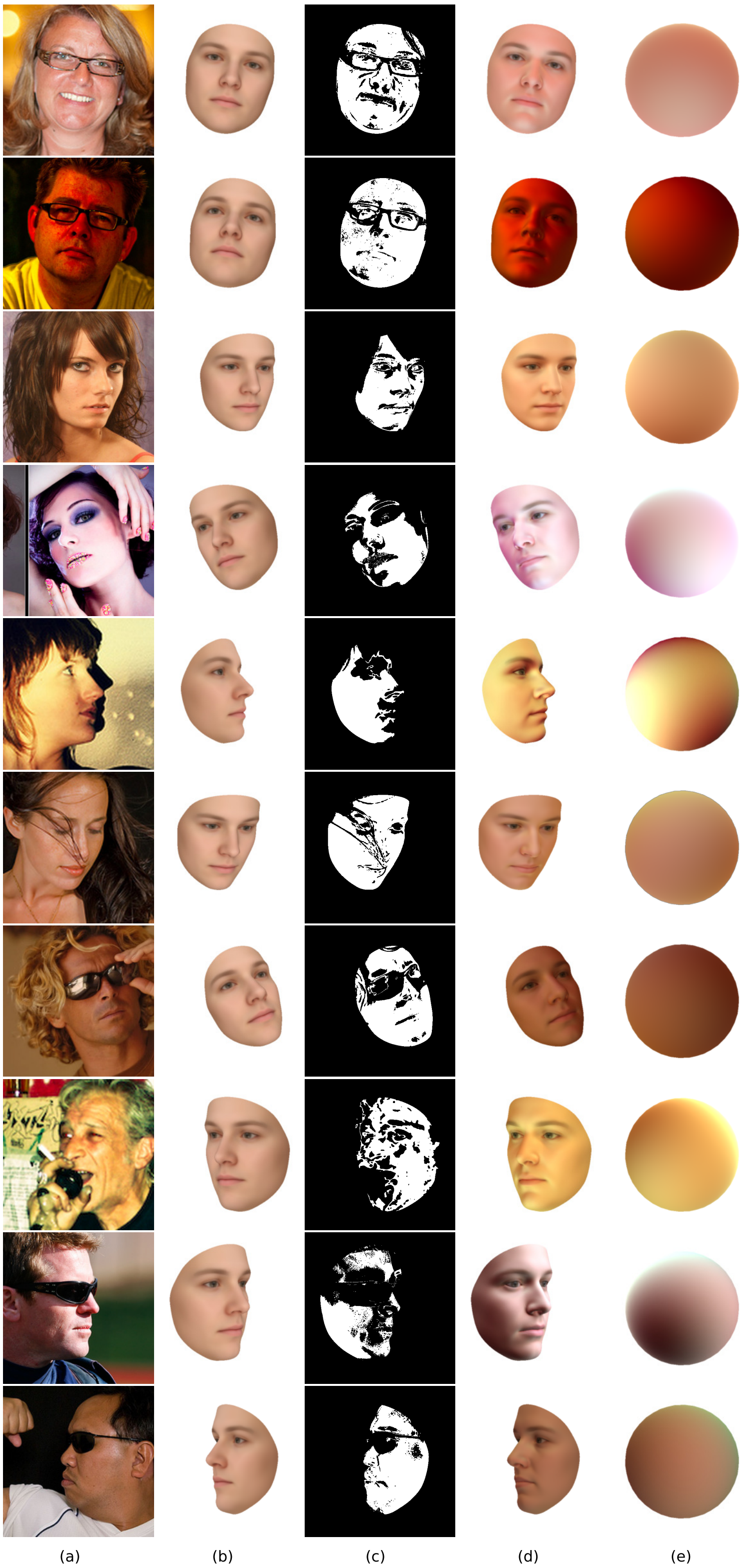

Fig. 15: A qualitative evaluation of the illumination estimation on the AFLW database (Köstinger et al (2011)). We show the target image (a), pose initialization (b), the obtained inlier mask (c), the mean face of the Basel Face Model rendered with the estimated illumination condition (d) and a sphere with the average face albedo rendered with the estimated illumination condition (e). The sphere provides a normalized rendering of the illumination condition. We observe that glasses, facial hair and various other occlusions are excluded from the illumination estimation. At the same time, we see minor limitations: things that are not well explained by our simplified illumination model (like specular highlights or cast shadows) or strong variations in facial appearance (e.g. in the mouth, eye or eyebrow region). Affected regions do not mislead the illumination estimation but are excluded by our robust approach.

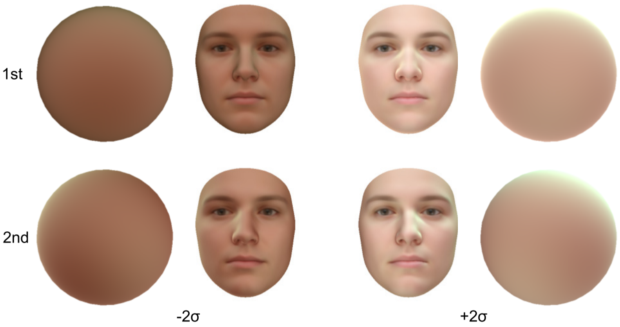

Fig. 16: We visualize the first two eigenmodes of our illumination prior using PCA. The first parameter represents luminosity. From the second eigenmode, we see that illuminations from above on both sides are very prominent in the dataset. The illumination conditions are rendered on the mean face of the Basel Face Model and a sphere with the average face albedo. From those eigenmodes we derive the strongest variation in illumination of natural images.

3.2 Occlusion-aware Model

For the 3DMM adaptation, the 3D pose has to be initialized. In the literature, this is performed manually (Blanz and Vetter (1999); Romdhani and Vetter (2003); Aldrian and Smith (2013)) or by using fiducial point detections (Schönborn et al (2013)). For our model adaptation experiments on the AR and LFW database, we use automated fiducial point detection results from the external CLandmark Library made available by Uřičář et al (2015). We integrate the fiducial point detections in our probabilistic framework as proposed in Schönborn et al (2013) and assume an uncertainty σ of 4 pixels on the detections. Our method is therefore fully automated and does not need manual input.

For the model adaptation experiments, we perform alternating 2,000 Metropolis-Hastings sampling steps (best sample is taken to proceed) followed by a segmentation step with 5 iterations and repeat this procedure five times. This amounts to a total of 10,000 samples and 25 segmentation iterations.

3.2.1 Segmentation

To evaluate the segmentation performance of the proposed algorithm, we manually segmented occlusions and beards in the AR face dataset proposed by Martinez and Benavente (1998). We took a subset of images consisting of the first 10 male and female participants appearing in both sessions. Unlike other evaluations, we include images under illuminations from the front and the side. We selected the neutral (containing glasses and beards) images from the first session and images with scarves and sunglasses from the second session. The total set consists of 120 images. We had to exclude 5 images because the fiducial point detection (4) or the illumination estimation (1) failed for them (m-009-25, w-009-25, w-002-22, w-012-24, w-012-25). For evaluation, we labeled an elliptical face region and occlusions manually. We added a separate label for beards to distinguish between occlusions and beards. The face region and occlusion inside it are labeled manually. The evaluation was done within the face region only. We have made our manual annotations used in this evaluation publicly available[5].

In our previous work (Egger et al (2016)), we compared our method to a standard technique to handle outliers, namely a trimmed estimator including only $n\%$ of the pixels which are best explained by the face model. In this work, we present the segmentation result including beards as an additional label. We present the simple matching coefficient (SMC) and the F1-Score for detailed analysis in Table 1. In our experiments, we distinguish the three image settings: neutral, scarf and sunglasses. We include the result of the initialization to depict its contribution and show that the fitting improves the segmentation even more.

3.2.2 Quality of Fit

We present qualitative results of our fitting quality on the AR face database (Martinez and Benavente (1998)) and the Labelled Faces in the Wild database (LFW) (Huang et al (2007)). In our results in Figure 18, the images include beards, occlusions, and un-occluded face images. In Figure 19 we also include results where our method fails. Our method detects occlusions by an appearance prior from the face model. If occlusions can be explained by the color or illumination model, the segmentation will be wrong.

Through the explicit modeling of beards, the errors in the beard region are reduced and the face model does not drift away anymore in this region (see Figure 21). An additional result of the inference process is an explicit beard type estimation. The quality of the model adaptation on the AFLW database shows almost the same performance as we obtain on data without occlusions. Interesting parts of the face are included gradually during the fitting process, see Figure 20. Our method also performs well on images without occlusions and does not tend to exclude parts of the face. In our recent work, we showed that our approach is not limited to neutral faces but it can also be applied using

[5] `http://gravis.cs.unibas.ch/publications/2017/2017_Occlusion-aware_3D_Morphable_Models.zip`

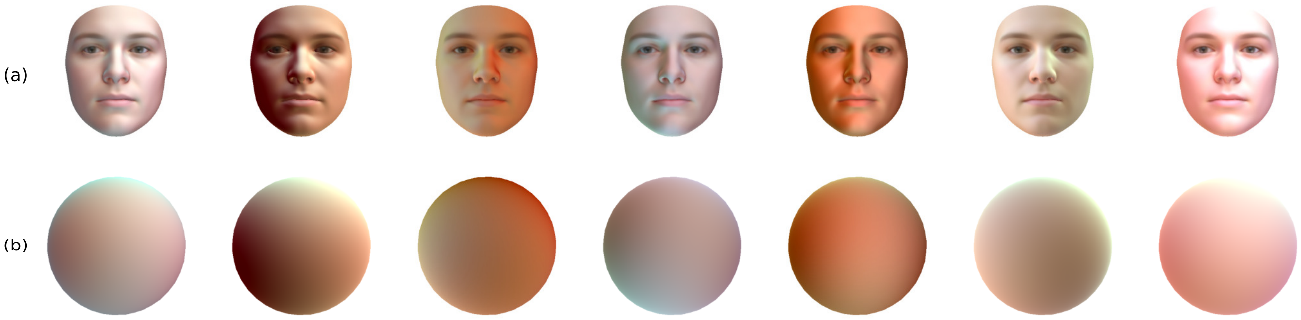

Fig. 17: Random samples from our illumination prior represent real-world illumination conditions. The proposed prior represents a wide range of different illumination conditions. The samples are rendered with the mean face of the Basel Face Model (a) and a sphere with the average face albedo (b).

Method	neutral	glasses	scarf
Initialisation z_{face}	0.78 (0.86|0.41)	0.81 (0.83|0.77)	0.73 (0.73|0.73)
Initialisation z_{beard}	0.97 (-|0.99)	0.95 (-|0.98)	1.00 (-|1.00)
Initialisation $z_{\text{non-face}}$	0.71 (0.09|0.83)	0.75 (0.67|0.80)	0.69 (0.66|0.69)
Full model z_{face}	0.84 (0.90|0.50)	0.85 (0.87|0.81)	0.85 (0.86|0.83)
Full model z_{beard}	0.98 (0.68|0.99)	0.96 (0.57|0.98)	0.98 (-|0.99)
Full model $z_{\text{non-face}}$	0.80 (0.14|0.89)	0.81 (0.72|0.86)	0.78 (0.59|0.85)

Table 1: Comparison of segmentation performance in SMC and in brackets the F1-Scores (class|rest) for all labels on the AR face database (Martinez and Benavente (1998)). We present separate results for our initialization using robust illumination estimation (line 1-3) and the full model including segmentation (line 4-6).

a model with facial expressions Egger et al (2017) like the Basel Face Model 2017 Gerig et al (2017).

3.2.3 Runtime

The method's performance in terms of speed is lower than that of optimization-only strategies. The stochastic sampler runs in approximately 20 minutes on current workstation hardware (Intel Xeon E5-2670 with 2.6 GHz), single-threaded. The segmentation takes around 2 minutes including the updates of the histograms in every step. The robust illumination estimation method needs 30 seconds.

The full inference process takes around 25 minutes per image. The long runtime is in large part due to the use of a software renderer and the high resolution of the face model, we did not invest in tuning computational performance. The overall computational effort is within the range of other techniques for full Morphable Model adaptation including texture and illumination.

3.2.4 Discussion

Quantitative evaluation of our full framework and a comparison to other state-of-the-art techniques is currently not possible. For in the wild images, especially under occlusion, there is no ground truth for the shape or other parameters like the illumination available. Such a benchmark would be highly beneficial for the community but is out of the scope of this publication. We choose to evaluate each component of our method in a quantitative manner and evaluated the full framework qualitatively. For future comparisons, we release our source code. In our previous work (Schönborn et al (2017)) we provide an in-depth performance analysis including a quantitative evaluation of 3D reconstruction. On these semi-synthetic renderings, we achieved state-of-the-art results.

4 Conclusion

We proposed a novel approach for combined segmentation and 3D Morphable Face Model adaptation. Jointly solving model adaptation and segmentation leads to an occlusion-aware face analysis framework and a semantic segmentation. Regions like the eye, eyebrow, nose, and mouth are harder to fit by the face model and are therefore often excluded by other robust methods. Our model includes those regions automatically during the inference process.

Illumination estimation is a critical part of robust image analysis. A proper initialization of the illumination parameters is necessary for the robust Analysis-

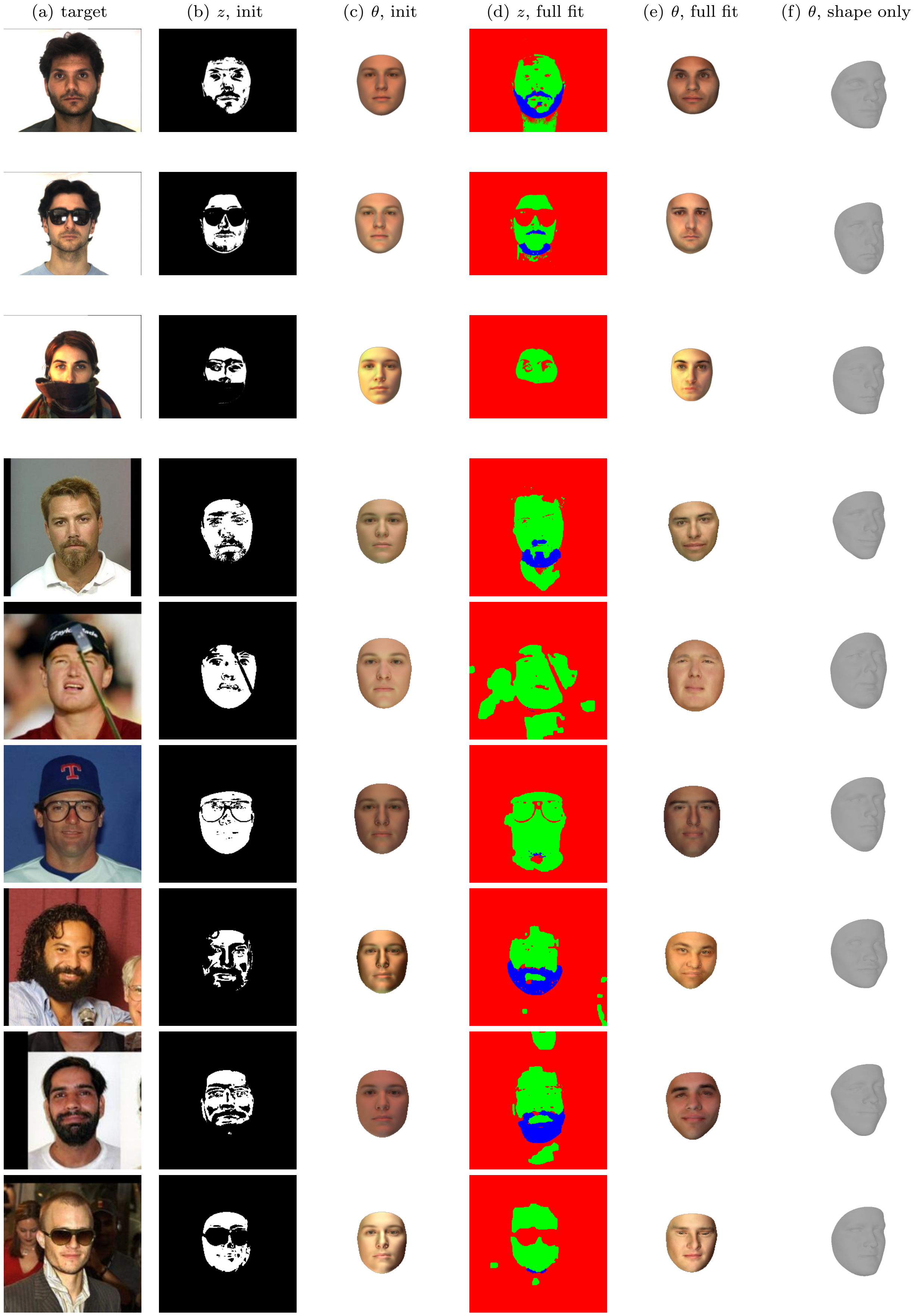

Fig. 18: (a) Target images from the AR face database (first three) Martinez and Benavente (1998) and the LFW database (Huang et al (2007)). (b) and (c) depict our initialisation arising from the robust illumination estimation, (d) and (e) present the final results. The shape component of our reconstruction is depicted in (f) under a normalized pose. Our final segmentation and synthesized face includes much more information of the eye, eyebrow, nose and mouth regions than the initialisation.

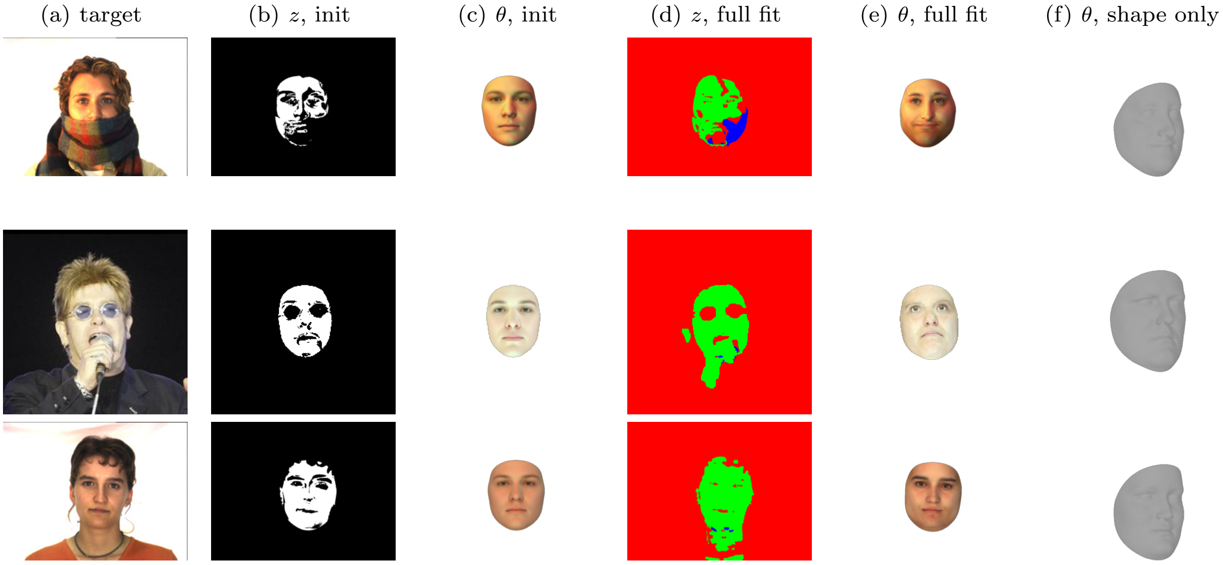

Fig. 19: We present usual cases of failure: scarves can be explained by the light and color model and are therefore mislabelled. Hands have similar color appearance and do not distort the face model adaptation but lead to a wrong segmentation. The prototype for chin-beards is mislead by shadows under the chin which are not modelled in our illumination model (such errors could be eliminated by more complex illumination models e.g. Schneider et al (2017)). Note that our method is not adapted to a specific kind of occlusion.

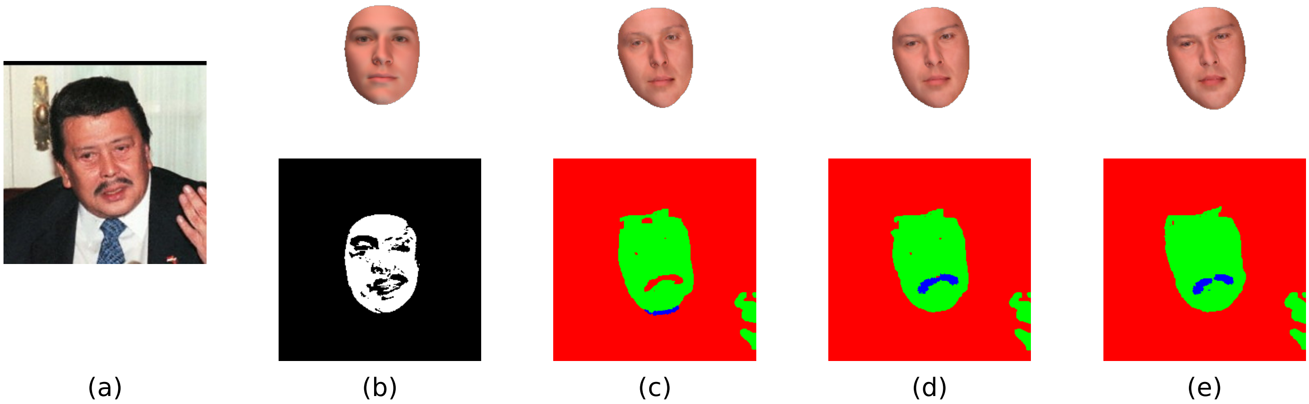

Fig. 20: The results at different steps in our framework including the EM-strategy. The target image is shown in (a). The result of the robust illumination estimation is shown in (b), we observe a strong pose misalignment in the roll angle. After the first 1000 samples of our model adaptation process the pose was adapted to the image (c) during the later model adaptation the correspondence gets better and the beard and face region are segmented better. We present the result after 3000 samples (d) and after the full 10’000 samples (e).

by-Synthesis process. We propose a RANSAC-like robust illumination estimation leading to good estimations even under occlusions.

Beards are especially challenging as they are not included in the face model and mislead the illumination estimation. We explicitly add a prototype-based beard model to overcome this limitation. Beards are segmented separately and guide the face model adaptation.

Additionally, we applied the robust illumination estimation technique on the AFLW database containing faces in a huge variety of scenes under arbitrary illumination conditions. The resulting prior is highly applicable for probabilistic frameworks as well as for data-greedy algorithms like deep learning methods for aug-

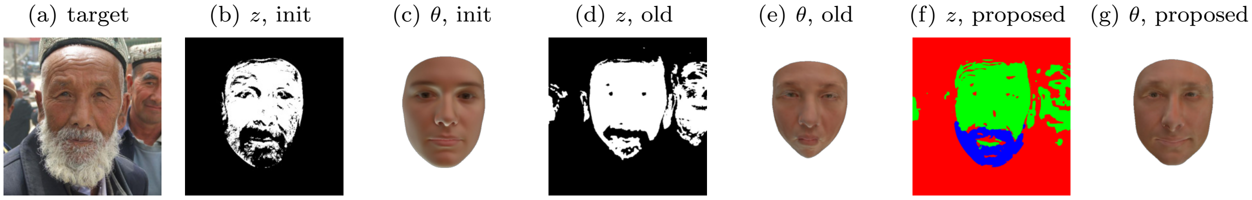

Fig. 21: We present the benefit of explicitly modeling beards. The beards cannot be reliably excluded during illumination estimation since they can partially be explained by illumination effects (b, c). We compare our results to the same approach without modeling beards explicitly (d, e, Egger et al (2016)). By explicitly modeling beards the face model adaptation is not misled by the beard region (f). Through the coupling of the beard model with the face model, the underlying face is kept at the correct position (g).

menting or generating data under unconstrained illumination conditions. The published illumination prior, from a broad range of real-world photographs, is the first, publicly available one.

Our model is a fully probabilistic occlusion-aware face analysis framework. It builds upon a generative model which is built from few high-quality face scans instead of a large amount of training data. It does not require any manual input or database adaptation and is therefore fully automated. The approach performs well on images under complex illumination settings and on an “in the wild” database for various kinds of occlusions.

References

Aldrian O, Smith WA (2013) Inverse rendering of faces with a 3d morphable model. IEEE Transactions on Pattern Analysis and Machine Intelligence 35(5):1080–1093

Arthur D, Vassilvitskii S (2007) k-means++: The advantages of careful seeding. In: Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms, Society for Industrial and Applied Mathematics, pp 1027–1035

Barron JT, Malik J (2015) Shape, illumination, and reflectance from shading. IEEE transactions on pattern analysis and machine intelligence 37(8):1670–1687

Basri R, Jacobs DW (2003) Lambertian reflectance and linear subspaces. Pattern Analysis and Machine Intelligence, IEEE Transactions on 25(2):218–233

Blanz V, Vetter T (1999) A morphable model for the synthesis of 3d faces. In: SIGGRAPH’99 Proceedings of the 26th annual conference on Computer graphics and interactive techniques, ACM Press, pp 187–194

Chan TF, Vese LA (2001) Active contours without edges. Image processing, IEEE transactions on 10(2):266–277

Dalca AV, Sridharan R, Cloonan L, Fitzpatrick KM, Kanakis A, Furie KL, Rosand J, Wu O, Sabuncu M, Rost NS, et al (2014) Segmentation of cerebrovascular pathologies in stroke patients with spatial and shape priors. In: Medical image computing and computer-assisted intervention: MICCAI International Conference on Medical Image Computing and Computer-Assisted Intervention, NIH Public Access, vol 17, p 773

De Smet M, Fransens R, Van Gool L (2006) A generalized em approach for 3d model based face recognition under occlusions. In: Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on, IEEE, vol 2, pp 1423–1430

Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via the em algorithm. Journal of the royal statistical society Series B (methodological) pp 1–38

Egger B (2017) Semantic morphable models. PhD thesis, PhD Thesis, University of Basel

Egger B, Schönborn S, Forster A, Vetter T (2014) Pose normalization for eye gaze estimation and facial attribute description from still images. In: German Conference on Pattern Recognition, Springer, pp 317–327

Egger B, Schneider A, Blumer C, Forster A, Schönborn S, Vetter T (2016) Occlusion-aware 3d morphable face models. In: British Machine Vision Conference (BMVC)

Egger B, Schönborn S, Blumer C, Vetter T (2017) Probabilistic morphable models. Statistical Shape and Deformation Analysis: Methods, Implementation and Applications p 115

Fischler MA, Bolles RC (1981) Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Communications of the ACM 24(6):381–395

Gerig T, Morel-Forster A, Blumer C, Egger B, Lüthi M, Schönborn S, Vetter T (2017) Morphable face models-an open framework. arXiv preprint arXiv:170908398

Gross R, Matthews I, Cohn J, Kanade T, Baker S (2010) Multi-PIE. Image and Vision Computing 28(5):807–813

Huang GB, Ramesh M, Berg T, Learned-Miller E (2007) Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Tech. Rep. 07-49, University of Massachusetts, Amherst

Huang R, Pavlovic V, Metaxas DN (2004) A graphical model framework for coupling mrfs and deformable models. In: Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on, IEEE, vol 2, pp II–739

Huber P, Feng ZH, Christmas W, Kittler J, Rätsch M (2015) Fitting 3d morphable face models using local features. In: Image Processing (ICIP), 2015 IEEE International Conference on, IEEE, pp 1195–1199

Jourabloo A, Liu X (2016) Large-pose face alignment via cnn-based dense 3d model fitting. In: CVPR

Kortylewski A (2017) Model-based image analysis for forensic shoe print recognition. PhD Thesis

Kortylewski A, Egger B, Schneider A, Gerig T, Forster A, Vetter T (2017) Empirically analyzing the effect of dataset biases on deep face recognition systems. arXiv preprint arXiv:171201619

Köstinger M, Wohlhart P, Roth PM, Bischof H (2011) Annotated facial landmarks in the wild: A large-scale, real-world database for facial landmark localization. In: Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on, pp 2144–2151

Kulkarni TD, Kohli P, Tenenbaum JB, Mansinghka V (2015) Picture: A probabilistic programming language for scene perception. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp 4390–4399

Le THN, Luu K, Savvides M (2015) Fast and robust self-training beard/moustache detection and segmentation. In: Biometrics (ICB), 2015 International Conference on, IEEE, pp 507–512

Lüthi M, Blanc R, Albrecht T, Gass T, Goksel O, Buchler P, Kistler M, Bousleiman H, Reyes M, Cattin PC, others (2012) Statismo-a framework for PCA based statistical models. The Insight Journal pp 1–18

Maninchedda F, Häne C, Jacquet B, Delaunoy A, Pollefeys M (2016) Semantic 3d reconstruction of heads. In: European Conference on Computer Vision, Springer, pp 667–683

Marschner SR, Greenberg DP (1997) Inverse lighting for photography. In: Color and Imaging Conference, Society for Imaging Science and Technology, vol 1997, pp 262–265

Martinez AM, Benavente R (1998) The ar face database. CVC Technical Report 24

Morel-Forster A (2017) Generative shape and image analysis by combining gaussian processes and mcmc sampling

Murphy KP, Weiss Y, Jordan MI (1999) Loopy belief propagation for approximate inference: An empirical study. In: Proceedings of the Fifteenth conference on Uncertainty in artificial intelligence, Morgan Kaufmann Publishers Inc., pp 467–475

Murphy-Chutorian E, Trivedi MM (2009) Head pose estimation in computer vision: A survey. IEEE transactions on pattern analysis and machine intelligence 31(4):607–626

Nguyen MH, Lalonde JF, Efros AA, De la Torre F (2008) Image-based shaving. In: Computer Graphics Forum, Wiley Online Library, vol 27, pp 627–635

Paysan P, Knothe R, Amberg B, Romdhani S, Vetter T (2009) A 3d face model for pose and illumination invariant face recognition. In: Proceedings of the 6th IEEE International Conference on Advanced Video and Signal based Surveillance (AVSS), IEEE, pp 296–301

Pierrard JS, Vetter T (2007) Skin detail analysis for face recognition. In: Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on, IEEE, pp 1–8

Ramamoorthi R, Hanrahan P (2001) An efficient representation for irradiance environment maps. In: Proceedings of the 28th annual conference on Computer graphics and interactive techniques, ACM, pp 497–500

Richardson E, Sela M, Kimmel R (2016) 3d face reconstruction by learning from synthetic data. arXiv preprint arXiv:160904387

Romdhani S, Vetter T (2003) Efficient, robust and accurate fitting of a 3d morphable model. In: Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on, IEEE, pp 59–66

Saito S, Li T, Li H (2016) Real-time facial segmentation and performance capture from rgb input. In: Leibe B, Matas J, Sebe N, Welling M (eds) Computer Vision – ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part VIII, Springer International Publishing, Cham, pp 244–261

Schneider A, Schönborn S, Egger B, Frobeen L, Vetter T (2017) Efficient global illumination for morphable models. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp 3865–3873

Schönborn S, Forster A, Egger B, Vetter T (2013) A monte carlo strategy to integrate detection and model-based face analysis. In: Pattern Recognition, Springer, pp 101–110

Schönborn S, Egger B, Morel-Forster A, Vetter T (2017) Markov chain monte carlo for automated face image analysis. vol 123, pp 160–183

Shahlaei D, Blanz V (2015) Realistic inverse lighting from a single 2d image of a face, taken under unknown and complex lighting. In: Automatic Face and Gesture Recognition (FG), 2015 11th IEEE International Conference and Workshops on, IEEE, vol 1, pp 1–8

Tewari A, Zollhöfer M, Kim H, Garrido P, Bernard F, Pérez P, Theobalt C (2017) Mofa: Model-based deep convolutional face autoencoder for unsupervised monocular reconstruction. arXiv preprint arXiv:170310580

Tu Z, Chen X, Yuille AL, Zhu SC (2005) Image parsing: Unifying segmentation, detection, and recognition. International Journal of computer vision 63(2):113–140

Uřičář M, Franc V, Thomas D, Akihiro S, Hlaváč V (2015) Real-time Multi-view Facial Landmark Detector Learned by the Structured Output SVM. In: 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), 2015, vol 02, pp 1–8

Wang Y, Liu Z, Hua G, Wen Z, Zhang Z, Samaras D (2007) Face re-lighting from a single image under harsh lighting conditions. In: Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on, IEEE, pp 1–8

Yildirim I, Janner M, Belledonne M, Wallraven C, Freiwald WA, Tenenbaum JB (2017) Causal and compositional generative models in online perception. In: to be published at 39th Annual Conference of the Cognitive Science Society

Zhu X, Yan J, Yi D, Lei Z, Li S (2015) Discriminative 3d morphable model fitting. In: Proceedings of 11th IEEE International Conference on Automatic Face and Gesture Recognition FG2015, Ljubljana, Slovenia

Zhu X, Lei Z, Liu X, Shi H, Li SZ (2016) Face alignment across large poses: A 3d solution. In: CVPR

Zivanov J, Forster A, Schönborn S, Vetter T (2013) Human face shape analysis under spherical harmonics illumination considering self occlusion. In: ICB-2013, 6th International Conference on Biometrics, Madrid