Statistically Learned Deformable Eye Models

Joan Alabort-i-Medina(✉), Bingqing Qu, and Stefanos Zafeiriou

Imperial College London, London, UK
{ja310,s.zafeiriou}@imperial.ac.uk, sylar.qu@gmail.com

Abstract. In this paper we study the feasibility of using standard deformable model fitting techniques to accurately track the deformation and motion of the human eye. To this end, we propose two highly detailed shape annotation schemes (open and close eyes), with +30 feature landmark points, high resolution eye images. We build extremely detailed Active Appearance Models (AAM), Constrained Local Models (CLM) and Supervised Descent Method (SDM) models of the human eye and report preliminary experiments comparing the relative performance of the previous techniques on the problem of eye alignment.

Keywords: Eye alignment · Eye tracking · Active appearance models · Constrained local models · Supervised descent method

1 Introduction

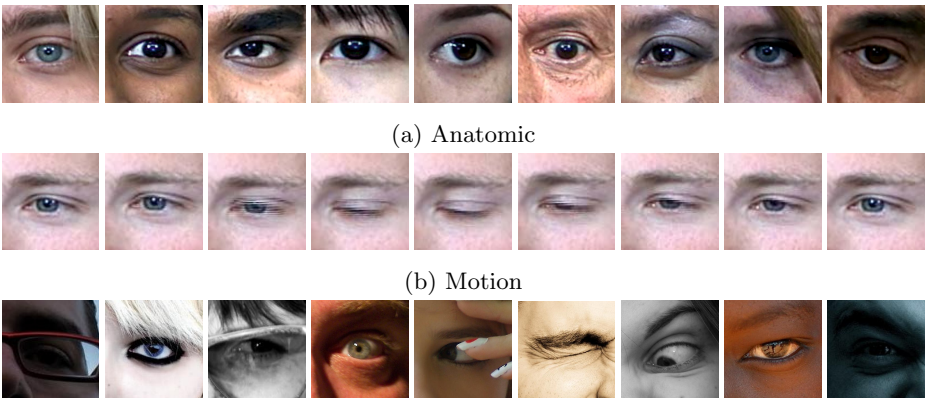
In recent years, the automatic analysis of facial images and video has attracted a lot of interest from the computer vision and machine learning research communities [18].

Within this context, eyes have proven to be among the most discriminative regions of the human face providing, for example, a reliable source of biometric information for face identification and recognition. On the same page, psychologist have reported strong evidence that the behaviour and movement of the eyes has strong connections with the brain cognitive processes [6] and offer important cues to understand the subtleness of facial behaviour [2]. On the other hand, gaze tracking is known to play an important role in the design of successful applications in human-computer interaction [10].

Consequently, the development of generic eye alignment algorithms capable of localizing and discriminating between different eye regions and capable of accurately describing the deformation and motion of the eyes is essential for the development of future human-centred-interfaces. For example, effective and reliable eye alignment is typically the first step in any deception and concealment-of-intent detection systems due to the proven correlation between eyelid movement and intentional deceit [5].

However, despite recent advances [10,11,14–16], accurate and robust eye alignment in unconstrained scenarios remains an extremely challenging task. The main difficulty arises from the very diverse appearance of eyes caused by both

L. Agapito et al. (Eds.): ECCV 2014 Workshops, Part I, LNCS 8925, pp. 285–295, 2015.
DOI: 10.1007/978-3-319-16178-5_19



(a) Anatomic

(b) Motion

(c) In-the-wild

Fig. 1. Appearance variability of eyes

anatomical differences between individuals (Figure 1a) and the high deformability and fast movement of the different eye components (Figure 1b). Moreover, other factors such as different illumination conditions, head pose and partial occlusion contribute to increase the appearance variability of the eyes in in-the-wild images (Figure 1c).

In this paper, we study the feasibility of using standard deformable model fitting techniques, such as Active Appearance Models (AAM) [4,8] and Constrained Local Models (CLM) [13], as well as the recently proposed Supervised Descent Method (SDM) [17], to accurately track the deformation and motion of the human eye. To this end, we propose two highly detailed shape annotation schemes (open and close eyes), with +30 feature landmark points, for annotating high resolution eye images. Using the previous schemes, we build extremely detailed eye models and conduct a preliminary study comparing the performance of the previous three techniques on the problem of eye alignment.

The remainder of the paper is structured as follows. Section 2 reviews prior work on eye tracking. Our newly proposed annotation schemes for open and close eyes are describe in Section 3. Section 4 offers a quick overview of the different deformable model fitting techniques considered in the paper, i.e. AAM, CLM and SDM. Experimental results are presented in Section 5 and conclusions are drawn in Section 6

2 Prior Work

A largely diverse number of approaches, ranging from simple techniques based on the application of edge detectors and Hough transform [14,16] to more sophisticated model-based approaches [10,11], have been used to solve the eye alignment problem.

The two closest works to the approach present in this paper are the ones of Moriyama et al. [10] and Orozco et al. [11]. The authors of [10] propose a 2D

handly-crafted parametrised generative eye shape model inspired by the anatomical structure of the human eye. Their approach requires manual initialization for the eye's texture. Fitting the previous eye model to a novel image is posed as an image alignment problem within the standard Lucas Kanade framework.

On the other hand, the authors of [11] propose an on-line appearance-based tracker that automatically adapts to changes in eye texture. They use a parametrised shape model based on a hand-crafted standard designed by the computer animation industry (Face Animation Parameters (FAP)). Fitting of their eye model is posed as a gradient descent optimization problem. Their approach requires careful manual initialization to ensure that the model gradually learns a useful representation of the eye texture.

Conversely, the models used in this paper make less assumption with respect to the shape and texture of the eyes since both of them are (either explicitly or implicitly) statistically learned from training data. Moreover, they can be automatically initialize using the coarse initialization provided by an off-the-shelf eye detector, removing the need for manual initialization. On the other hand, these thechniques rely on the availability of annotated training data.

3 Eye Model

Eyes are highly deformable organs that can be decomposed in several different parts [10] (Figure 3). Some of this parts might become partially or completely occluded by others due to the natural motion of the eyes. For example, on the open right eye images in Figure 1b all five different regions: upper lid, lower lid, sclera, iris and pupil are visible. In contrast, on the half-open and closed right eye images on the same figure, only the some of the previous parts are visible and the rest are naturally occluded.

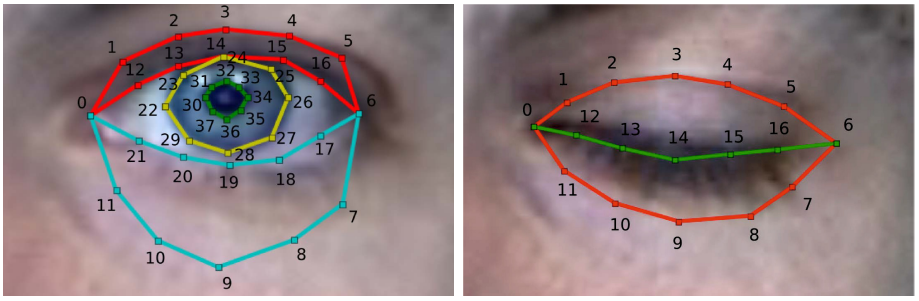
In order to fit eyes using standard deformable model fitting techniques, one needs to define the shape of the object being modeled explicitly, as a set of feature landmark points. While this might be simple for some objects (e.g. frontal faces or rear cars), the self-occluding nature of the eyes produces drastic changes in their appearance making the definition of a single set of feature landmark points non trivial.

In this work, we propose to solve the previous problem by differentiating between full/half-open eyes and close eye and use two different sets of landmarks to describe the shape of the eyes in both of these states. Note that, although the open/half-open eye landmarks are adequate to describe most of the eye motion they cannot deal with the singularity that a close eye represents (it would be indeed very difficult to annotate close eye images using the set of landmark points describing the shape of full/half-open eyes).

A direct consequence of the previous decision is that all deformable model fitting techniques will need to differentiate between full/half-open eyes and close eyes. Hence, given a novel eye image this techniques will need to fit both full/half-open eye and close eye models to the image and evaluate the correctness of each model using a particular score metric. In this paper, we use a simple Support Vector Machine (SVM) classifier to determining the correctness of each model.

(a) Upper eyelid (b) Lower eyelid (c) Sclera (d) Iris (e) Pupil

Fig. 2. A possible decomposition of eyes in different parts



(a) Full/half-open eye (b) Close eye

Fig. 3. Full/half-open and close eye feature landmarks points

Full/Half-Open Eye Model. The shape of full/half-open eyes is described by set of 38 feature landmarks points annotating the five different eye regions depict in Figure 3, i.e. (i) upper eyelid, (ii) lower eyelid, (iii) sclera, (iv) iris and, (v) pupil. A detailed diagram with the specific meaning of each landmark is shown in Figure 3a.

Close Eye Model. To describe the shape of close eyes 17 feature landmark points are used. Note that, the upper and lower eyelid are the only parts visible in this state. An annotated close eye is shown in Figure 3b.

4 Deformable Eye Fitting

This section reviews the three different deformable model fitting techniques used in this paper, i.e. AAM, CLM and SDM.

4.1 Active Appearance Models

Active Appearance Models (AAM) [4,8] are global deformable models that describe the shape and texture of a particular object as a linear combination of a set of bases. The shape model is built from a set of landmarks describing the shape of the object. These landmarks are first normalized with respect to a

2D global similarity transform and then Principal Component Analysis (PCA) is applied to obtained a set of linear shape bases. The previous basis are composed with a 2D global similarity transform that allows shapes to be arbitrarily positioned on the image coordinate system. The shape model can be mathematically expressed as:

$$\mathbf{s} = s\mathbf{R}(\bar{\mathbf{s}} + \mathbf{V}\mathbf{p}) + \mathbf{t}_{\mathbf{x},\mathbf{y}} \tag{1}$$

where $\bar{\mathbf{s}} \in \mathcal{R}^{2v\times 1}$ is the mean shape, and $\mathbf{V} \in \mathcal{R}^{2v\times n}$ and $\mathbf{p} \in \mathcal{R}^{n\times 1}$ denote the shape eigenvectors and shape parameters, respectively. Note that, s, $\mathbf{R}$ and $\mathbf{t}_{\mathbf{x},\mathbf{y}}$ contain the scale, rotation and translation parameters of the 2D global similarity transform.

The AAM's texture model is obtained by warping the texture information onto a common reference frame (generating the so called shape-free textures) and applying PCA to the vectorized warped textures. The texture model is defined by the following expression:

$$\mathbf{t} = \bar{\mathbf{t}} + \mathbf{U}\mathbf{c} \tag{2}$$

where $\mathbf{t} \in \mathcal{R}^{F\times 1}$ is the mean texture, and $\mathbf{U} \in \mathcal{R}^{F\times m}$ and $\mathbf{c} \in \mathcal{R}^{m\times 1}$ denote the texture eigenvectors and texture parameters, respectively.

Figure 4 and Figure 5 show the mean and first three principal components of a full/half-open open and a close eye intensity-based AAM using the annotation scheme described in the previous section.

Fitting Active Appearance Models (AAM). Fitting an AAM consists of minimizing the Sum of Squared Differences (SSD) between a vectorized warped image (given a first estimate of the shape, the image is warped onto the reference frame and then vectorized) and the linear texture model:

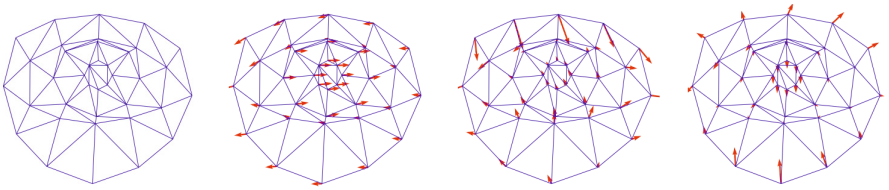
$$\mathbf{p}_o, \mathbf{c}_o = \arg\min_{\mathbf{p},\mathbf{c}} ||\mathbf{p}||^2_{\mathbf{\Lambda}^{-1}} + ||\mathbf{c}||^2_{\mathbf{\Sigma}^{-1}} + \frac{1}{\sigma^2}||\mathbf{i}[\mathbf{p}] - \bar{\mathbf{t}} + \mathbf{U}\mathbf{c}||^2 \tag{3}$$

Where $\mathbf{i}[\mathbf{p}] = \text{vec}(\mathcal{I} \circ \mathcal{W}(\mathbf{p}))$ denotes the vectorized warped image, $\mathbf{\Lambda}$ and $\mathbf{\Sigma}$ are diagonal matrices containing the eigenvalues associated to the shape and texture eigenvectors $\mathbf{V}$ and $\mathbf{U}$ respectively, and σ^2 quantifies the estimated uncertainty about image.

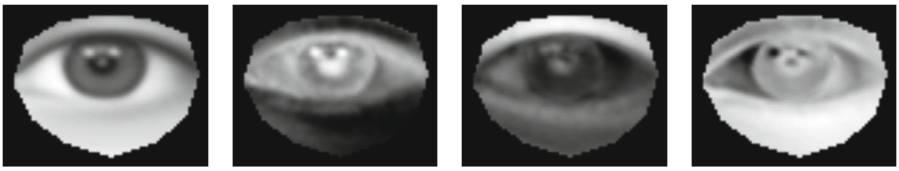
There exist several algorithms to solve the previous optimization problem [1,4,8,12]. A concise review can be found in [9]. In these paper, we use the Alternating Inverse Compositional (AIC) algorithm proposed by the authors of [12]. For further details on AAM and the AIC algorithm the reader is referred to [12] and [9].

4.2 Constrained Local Models (CLM)

Constrained Local Models (CLM) [3,13] are parts-based deformable models that define the texture of a particular object as independent local image regions

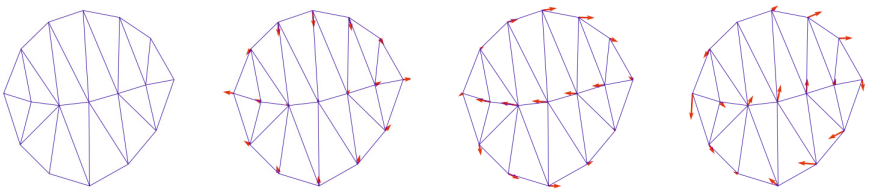


(a) Shape model

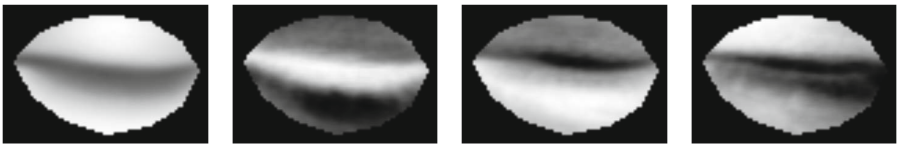


(b) Texture model

Fig. 4. Full/half-open eye Active Appearance Model



(a) Shape model



(b) Texture model

Fig. 5. Close eye Active Appearance Model

around each landmark. Their shape is represented using the same global PCA-based shape model used by AAM.

Even though generative approaches could be used to model local image regions, the usual approach is discriminative. For each landmark a classifier that quantifies the likelihood of the landmark being correctly aligned is learned based on the support of its local image region. The previous likelihood can be defined as:

$$\ell(l_i = 1|\mathbf{x}_i, \mathcal{I}) = \frac{1}{1 + \exp\{l_i \, \mathcal{C}_i(\mathcal{I}, \mathbf{x}_i)\}} \tag{4}$$

where $\mathcal{C}_i$ denotes a linear classifier that discriminates between aligned and misaligned locations, i.e.:

$$\mathcal{C}_i(\mathcal{I}, \mathbf{x}_i) = \mathbf{w}_i \left[\mathcal{I}(\mathbf{y}_i), \cdots, \mathcal{I}(\mathbf{y}_m)\right] + b_i \tag{5}$$

and $\{\mathbf{y}_i\}_{i=1}^{m} \in \mathbf{\Omega}_{\mathbf{x}_i}$, i.e. the image patch around the current landmark estimate $\mathbf{x}_i$.

Fitting Constrained Local Models. Fitting Constrained Local Models involves solving the following optimization problem [13]:

$$\mathbf{p}_o = \arg\min_{\mathbf{p}} ||\mathbf{p}||^2_{\mathbf{\Lambda}^{-1}} + \sum_{\mathbf{x}_i \in \mathbf{s}} \sum_{j=1}^{K} \frac{w_j}{\rho^2} ||\mathbf{x}_i - \mathbf{y}_j||^2 \tag{6}$$

where $w_j = \frac{1}{1+\exp\{l_i\, \mathcal{C}_j(\mathcal{I}, \mathbf{x}_j)\}}$ denotes the likelihood of each candidate landmark $\mathbf{y}_j$ in a particular local patch, $\mathbf{\Lambda}$ is a diagonal matrix containing the eigenvalues associated to the eigenvectors $\mathbf{V}$ of the shape model and ρ^2 quantifies the estimated uncertainty about shape.

The previous optimization problem can be solve using several strategies[3,13]. See [13] for a detailed review. The most popular technique for solving the expression in Equation 6 is the Regularised Landmark Mean-Shift (RLMS) algorithm proposes by Saragih et al. in [13]. This is the approach used in this paper. For further details on CLM and the RLMS the reader is referred to [13].

4.3 Supervised Descent Method (SDM)

The Supervised Descent Method (SDM) [17] is a recently proposed techniques for solving general nonlinear optimisation problems in computer vision. This technique can be used to solve the deformable model fitting problem by defining a local appearance model around each landmark (similar to the one defined by CLM) and an implicit non-parametric shape model.

SDM is posed as the cascade regression problem in which the following expression is optimised at each level:

$$\mathbf{R}_o^k, \mathbf{b}_o^k = \arg\min_{\mathbf{R}^k, \mathbf{b}^k} \sum_{i=1}^{N} \sum_{j=1}^{M} ||\mathbf{s}_{i,*} - \mathbf{s}_{i,j}^k + \mathbf{R}^k \Phi(I_i, \mathbf{s}_{i,j}) + \mathbf{b}^k||_2^2 \tag{7}$$

Where N and M index the total number of images and perturbation respectively, $\mathbf{s}_{i,*} \in \mathcal{R}^{2v \times 1}$ denotes the correct position of the shape landmarks in a particular image I_i, $\mathbf{s}_{i,j}$ are the perturbed version of $\mathbf{s}_{i,*}$ that we wish to correct, $\Phi(I_i, \mathbf{s}_{i,j})$ denotes the vectorized features extracted at each local appearance region, and finally $\mathbf{R}_o^k$ and $\mathbf{b}_o^k$ are, respectively, the regression matrix and bias term that minimise the previous expression.

The solutions $\mathbf{R}_o^k$ and $\mathbf{b}_o^k$ are obtained in closed-form by solving a linear least squares problem at each cascade level. Once inferred, $\mathbf{R}_o^k$ and $\mathbf{b}_o^k$ are used

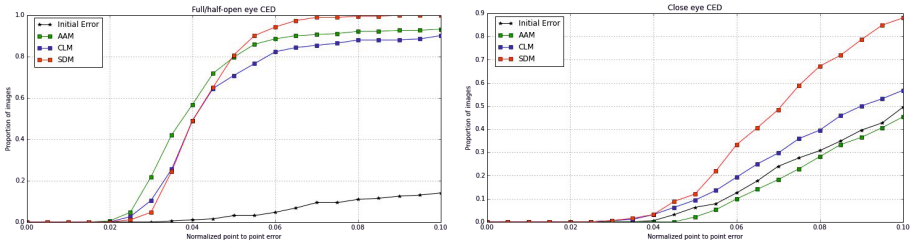


Fig. 6. CED curves for full/half-open and close eyes, respectively

to correct the position of $\mathbf{s}_{i,j}^{k}$ generating $\mathbf{s}_{i,j}^{k+1}$ and, with it, the next regression level of the cascade. In our implementation, this approach typically converges after 4 or 5 cascade levels. For more details on the SDM problem formulation for deformable object fitting and a more detailed explanation of its solution the reader is referred to [17].

5 Experiments

This section reports the performance of the previous three deformable model fitting techniques on the problems of eye alignment and eye tracking.

We report results for two different experiments. The first one compares, quantitatively, the accuracy of each technique, i.e. AAM, CLM, SDM. The second experiment shows qualitative results on the Helen [7] dataset, a recently proposed facial dataset containing high resolution in-the-wild images.

Note that, in order to save valuable space, all results are reported only for right eye models. We empirically verified that the results for right and left eye models are statistically equivalent (which is expected due to the obvious symmetry of the eyes).

5.1 Quantitative Eye Alignment Results

We start by evaluating the relative performance of each method on the problem of eye alignment.

For this experiment we collected and annotated two small dataset of 400 high resolution eye images each; one containing full/half-open eye images and the other close eye images. We randomly divide the available 400 annotated full/half-open eye images into equally sized training and testing sets. We train each model with the previous training set and report the accuracy their fitting accuracy on the testing set. The procedure is repeated for the available 400 annotated images containing closed eyes. Accuracy is reported using the error measure defined in [19], in which "face size" is simply replaced by the analogous "eye size". All methods are initialized by randomly perturbing the correct similarity transform and applying it to the mean shape of each model. Exemplar initializations obtained using the previous procedure are display in Figure 7.

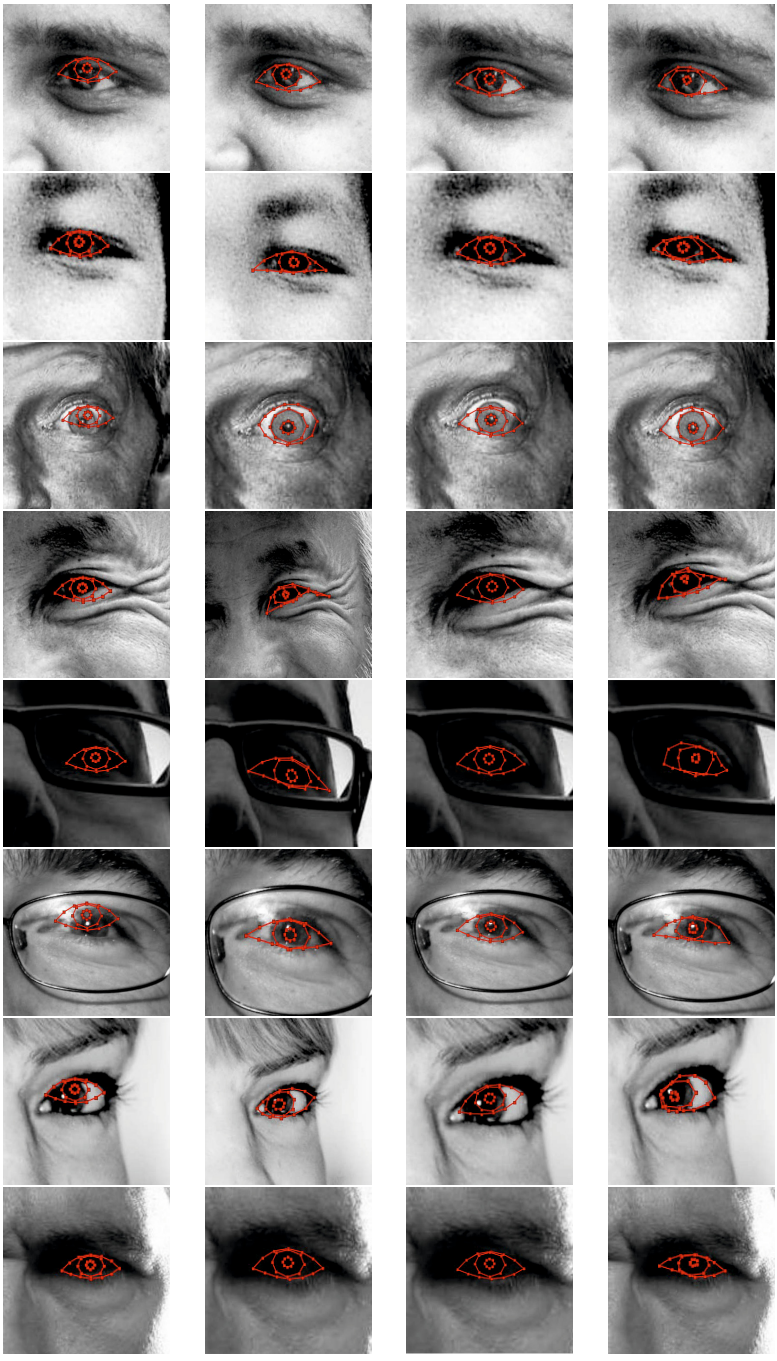


Fig. 7. Qualitative results on the Helen dataset. Columns for each subjects show: initialization, AAM result, CLM result and SDM result, respectively.

Figure 6 shows the Cumulative Error Distribution (CED) curves for full/half-open and close eyes. The results show that all methods are significantly more accurate fitting full/half-open eyes than close eyes. In particular, for full/half-open eyes, AAM is the most accurate method (it obtains the best results in the significant region $0.425 < err < 0.045$) while SDM is the most robust (it approaximatly fits all images with $err < 0.065$). CLM appear to be consistently inferior to AAM and SDM. It is worth noting, that all methods are capable of fitting the sclera, iris and pupil parts accurately and the reported errors are driven by the top (landmarks: [1-5]) and bottom landmarks (landmarks: [7-11]) of the upper and lower eyelids.

SDM is the most performant method for close eyes. The poor accuracy of all methods fitting close eyes images can be explained by the lack of meaningful features that can be extracted from the annotated close eye images, Figure 3b and Figure 5. This suggest that more contextual information (from the eyebrow or nose regions) might be necessary to accurately track close eyes using the previous methods.

5.2 Eye alignment in-the-wild

This experiment reports qualitative eye fitting results on images from the Helen dataset. All methods where initialized using the exact same procedure described in the previous experiment. Results for the three different techniques are shown in Figure 7.

6 Conclusions

In this paper we study the use of statistically learned models for deformable eye fitting. We introduce two novel shape annotation schemes, one for full/half-open eyes and another for close eyes, specifically designed to accurately annotate high resolution eye images. Finally, we report preliminary results comparing the performance of three different deformable model fitting techniques, i.e. Active Appearance Models, Constrained Local Models and Supervised Descent Method on the problem of eye alignment.

Acknowledgments. The work of Joan Alabort-i-Medina was funded by a DTA studentship from Imperial College London and by the Qualcomm Innovation Fellowship. The work of Stefanos Zafeiriou was partially funded by the EPSRC project EP/J017787/1 (4DFAB).

References

1. Amberg, B., Blake, A., Vetter, T.: On compositional image alignment, with an application to active appearance models. In: CVPR (2009)
2. Cohen, I., Sebe, N., Garg, A., Chen, L.S., Huang, T.S.: Facial expression recognition from video sequences: Temporal and static modeling. CVIU (2003)

3. Cootes, T.F., Taylor, C.J., Cooper, D.H., Graham, J.: Active shape models—their training and application. Computer Vision and Image Understanding (1995)
4. Cootes, T.F., Edwards, G.J., Taylor, C.J.: Active appearance models. IEEE Transactions on Pattern Analysis and Machine Intelligence (2001)
5. Fukuda, K.: Eye blinks: new indices for the detection of deception. International Journal of Psychophysiology (2001)
6. Stern, J.A., Walrath, L.C., Goldstein, R.: The endogenous eyeblink. Psychophysiology (1984)
7. Le, V., Brandt, J., Lin, Z., Bourdev, L., Huang, T.S.: Interactive facial feature localization. In: Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C. (eds.) ECCV 2012, Part III. LNCS, vol. 7574, pp. 679–692. Springer, Heidelberg (2012)
8. Matthews, I., Baker, S.: Active appearance models revisited. International Journal of Computer Vision (2004)
9. i Medina, J.A., Zafeiriou, S.: Bayesian active appearance models. In: CVPR (2014)
10. Moriyama, T., Kanade, T., Xiao, J., Cohn, J.F.: Meticulously detailed eye region model and its application to analysis of facial images. TPAMI (2006)
11. Orozco, J., Roca, X., Gonzlez, J.: Real-time gaze tracking with appearance-based models. Machine Vision and Applications (2009)
12. Papandreou, G., Maragos, P.: Adaptive and constrained algorithms for inverse compositional active appearance model fitting. In: CVPR (2008)
13. Saragih, J.M., Lucey, S., Cohn, J.F.: Deformable model fitting by regularized landmark mean-shift. International Journal of Computer Vision (2011)
14. Sirohey, S., Rosenfeld, A., Duric, Z.: A method of detecting and tracking irises and eyelids in video. Pattern Recognition (2002)
15. Tan, H., Zhang, Y.J.: Detecting eye blink states by tracking iris and eyelids (2006)
16. Wu, Y., Liu, H., Zha, H.: A new method of detecting human eyelids based on deformable templates. In: SMC (2004)
17. Xiong, X., De la Torre, F.: Supervised descent method and its application to face alignment. In: CVPR (2013)
18. Zeng, Z., Pantic, M., Roisman, G., Huang, T.: A survey of affect recognition methods: Audio, visual, and spontaneous expressions. TPAMI (2009)
19. Zhu, X., Ramanan, D.: Face detection, pose estimation, and landmark localization in the wild. In: CVPR (2012)