# Light-Invariant Fitting of Active Appearance Models

Daniel Pizarro
Alcalà University - Madrid

Julien Peyras
LASMEA - Clermont-Ferrand

Adrien Bartoli
LASMEA - Clermont-Ferrand

## Abstract

*This paper deals with shading and AAMs. Shading is created by lighting change. It can be of two types: self-shading and external shading. The effect of self-shading can be explicitly learned and handled by AAMs. This is not however possible for external shading, which is usually dealt with by robustifying the cost function.*

*We take a different approach: we measure the fitting cost in a so-called Light-Invariant space. This approach naturally handles self-shading and external shading. The framework is based on mild assumptions on the scene reflectance and the cameras. Some photometric camera response parameters are required. We propose to estimate these while fitting an existing color AAM in a photometric 'self-calibration' manner.*

*We report successful results with a face AAM with test images taken indoor under simple lighting change.*

## 1. Introduction

Active Appearance Models (AAMs) were introduced in [4] and since then have been the topic of many studies such as [2, 9]. They model the visual shape and appearance of an object or an object class and have been particularly successful to model faces. Face AAMs are used in this paper without loss of generality.

One of the main weaknesses of the AAMs is that there reliability generally degrades while the amount of variability increases in the training data. Sources of variability include person identity, expression, pose and shading. An AAM trained for several such sources will generally be prone to fall into local minima.

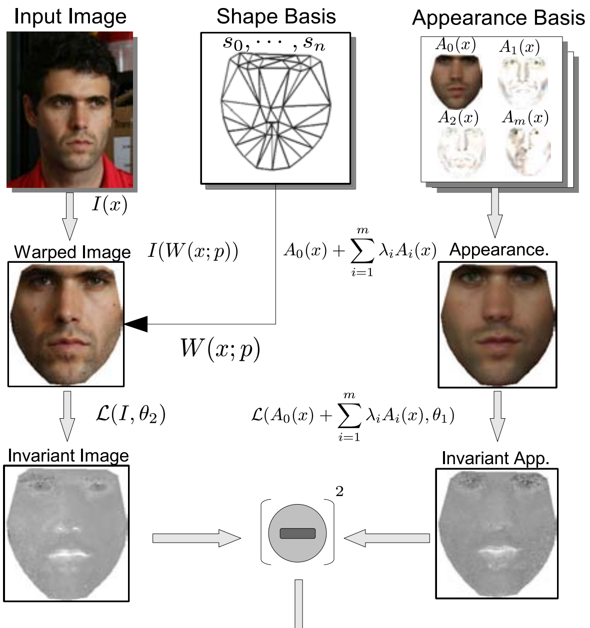
On the other hand, an AAM which has not been specifically trained to handle shading drifts away from the expected solution when the lighting changes. This makes AAMs useless for most of the applications where lighting conditions cannot be kept under control. Several recent papers consider this problem. Global illumination changes are modeled thanks to an affine normalization in the appearance basis [1]. Shaded areas are rejected as outliers thanks to a robust cost function [12]. These solutions are not fully satisfactory in an unconstrained lighting context.

Another approach is to build a rich training database including a large amount of lighting variations [11]. However, for many applications, we cannot reasonably pretend to be able to collect sufficiently many such training data. In addition, it is useful not to overload the appearance statistics but rather to separate the different sources of variability. Keeping the model complexity as low as possible usually results in improved fitting performances and better accuracy. A paper which recently followed this direction is [6]. It uses an Active Illumination Appearance (AIA) that models self-shading by additively combining two appearance bases, one for identity and one for self-shading.

We tackle the AAM fitting problem in the unconstrained illumination context. We do not explicitly model the appearance variations due to shading. This is based on the Light-Invariant transformation of [5]. From color training data acquired under any canonical illumination condition (which is also homogeneous in most cases) with a single camera, the *Light-Invariant AAM* fitting procedure we propose fits faces taken under uncontrolled illumination conditions. The only assumptions are a simplified model for the scene BRDF (Bidirectional Reflectance Distribution Function) and the photometric camera response. This is partly inspired by [10] in which direct Light-Invariant homography computation is successfully demonstrated.

Combining the Light-Invariant theory with AAMs not only allows us to efficiently fit AAMs to face images for which the lighting conditions are uncontrolled, it further allows synthesizing the canonical illumination appearance of the face. Thereafter, we assume that the canonical illumination is homogeneous. It has been shown that shadow free appearance reconstruction is in general an ill-posed problem [5]. We benefit from the strong prior knowledge that the observed object is known. We show successful shadow free face reconstruction. An overview of the proposed method is presented in Figure 1.

The paper is organized as follows. In §2 we describe the original AAM formulation and the Light-Invariant space. We discuss the possibility of combining the AAM frame-




$$\min_{\Phi} \sum_x \left[ \mathcal{L}(A_0(x) + \sum_{i=1}^{m} \lambda_i A_i(x), \theta_1) - \mathcal{L}(I(W(x;p)), \theta_2) \right]^2$$

$$\Phi = (p, \lambda_1, \cdots, \lambda_m, \theta_1, \theta_2)$$

Figure 1. Overview of the proposed method.

work with the Light-Invariant image theory in §3. Two algorithms are introduced. The basic Light-Invariant AAM (LI-AAM) fitting, with its mathematical development, are given in §4. The photometric 'self-calibration' of a regular AAM is explained in §5. We consider the possibility to handle different training and test cameras. This allows us to reconstruct a shadow free RGB appearance, as described in §6. In §7, we experimentally compare the fitting performance obtained with classical fitting of a greylevel AAM, a color AAM, and various LI-AAM fittings. We conclude and give our perspectives in §8.

# 2. Background and General Points

## 2.1. Active Appearance Models

An AAM combines two linear subspaces, one for the shape and one for the appearance. They are both learnt from a labelled set of training images [4].

### 2.1.1 The Shape Counterpart

The shape of an AAM is defined by the $N$ vertex coordinates of a mesh $s$ describing the object boundaries and deformations:

$$s = \left(u_1, v_1, u_2, v_2, \cdots, u_N, v_N\right), \tag{1}$$

where $u_i, v_i$ are the coordinates of vertex $i$. Principal Component Analysis (PCA) is applied to training shapes, centered on the mean shape $s_0$. A shape subspace $\boldsymbol{B}_s = [s_1, \cdots s_n]$ of $n-4$ shape eigencomponents is obtained, reducing the dimensionality of the training set. Four eigenvectors $s_{n-3}, \cdots, s_n$ are added to model 2D similarity transformations [9].

An instance of shape $s(p)$ is defined as a linear combination of shape eigenvectors with weights $p = (p_1, \cdots, p_n)$:

$$s(p) = s_0 + \sum_{i=1}^{n} p_i s_i. \tag{2}$$

Using the generated set of meshes parameterized by $p$, a warp function $W(x;p)$ is defined as a piecewise affine transformation from the base shape $s_0$ to the transformed mesh $s(p)$:

$$x' = W(x;p) \quad x = \begin{pmatrix} u \\ v \end{pmatrix} \quad x' = \begin{pmatrix} u' \\ v' \end{pmatrix}, \tag{3}$$

where $(u', v')^T$ is the warped coordinates.

### 2.1.2 The Appearance Counterpart

The training images are warped using $W(x;p)$ to normalize these in size. A photometric normalization is then applied to get the appearance training data. PCA is applied on these data in order to build a linear appearance subspace $\boldsymbol{B}_a = [A_1, \cdots, A_m]$ of $m-2$ image eigencomponents centered on the mean appearance image $A_0$. To compensate for bias and gain in the image intensity, the mean vector $A_0$ and a constant image $A_I$ are added (corresponding to the $m-1^{th}$ and $m^{th}$ components). This is the same process for color and greylevel images: the number of channels in the appearance images matches that of the training images.

An instance of appearance is defined as a linear combination of the $A_i(x)$ weighted by a set of parameters $\lambda = (\lambda_1, \cdots, \lambda_m)$:

$$A(x) = A_0(x) + \sum_{i=1}^{m} \lambda_i A_i(x). \tag{4}$$

### 2.1.3 Fitting

Fitting an AAM consists to find the shape and appearance parameters that make it best match the input image. A cost function is minimized with respect to the shape and appearance parameters:

$$\sum_x \left[ A_0(x) + \sum_{i=1}^{m} \lambda_i A_i(x) - I(W(x;p)) \right]^2. \tag{5}$$

The cost function (5) evaluates the pixel value discrepancy between the warped input image and an instance of appearance of the model. An iterative Gauss-Newton minimization process is used to retrieve parameters $p$ and $\lambda$.

### 2.2. Light-Invariant Image Theory

The transformation for Light-Invariant image formation is based on [5], which proposes a method for a single color image. This method relies on a simplified photometric image formation model where all surface materials follow a lambertian model, the lights are modeled as planckian sources and the camera sensor is narrowband. We consider faces to be lambertian to some extent. This assumption suffices for a range of lightings but might break down in cases such as specularities.

A more complex model of the camera is proposed in [8, 7], which takes into account nonlinearities in the image formation process (*i.e.* vignetting and the radiometric response function of the camera), and its usage for photometric alignment of images. Such a model allows to compensate for the camera response, which improves the applicability of a simplified image formation in general purpose cameras.

Given the three color components $\rho = (\rho_1, \rho_2, \rho_3)$, log chromaticity ratios are formed:

$$\mathcal{X}_1 = \log\left(\frac{\rho_1}{\rho_3}\right), \qquad \mathcal{X}_2 = \log\left(\frac{\rho_2}{\rho_3}\right). \tag{6}$$

An illumination invariant quantity $\mathcal{L}$ can be found by projecting $(\mathcal{X}_1, \mathcal{X}_2)$ along direction $\bar{e}^{\perp} = (\cos(\theta), \sin(\theta))$ parameterized by its angle $\theta$ from which:

$$\mathcal{L}(\rho, \theta) = \mathcal{X}_1(\rho)\cos(\theta) + \mathcal{X}_2(\rho)\sin(\theta). \tag{7}$$

This projection equates two colors corresponding to point viewed under different illuminations. It thus maps a color $\rho$ to its corresponding Light-Invariant representation.

Transforming the value of each pixel of the color image $\mathcal{S}$ results in $\mathcal{L}(\mathcal{S}, \theta)$, a 1D shadow invariant image. This image transformation is global in that it does not depend on the pixel position $q \in \mathbb{R}^2$, but only on its color value.

The only relevant parameter governing the transformation is the angle $\theta$ of the invariant line, which only depends on the camera spectral properties. We elevate the RGB data to a log, obtaining what we call the *logRGB space*. The transformation opportunely becomes linear:

$$\mathcal{L}(\rho, \theta) = L(\theta)\begin{pmatrix}\log(\rho_1)\\ \log(\rho_2)\\ \log(\rho_3)\end{pmatrix}, \tag{8}$$

with $L(\theta) = \big(cos(\theta), sin(\theta)\big)\begin{pmatrix}1 & 0 & -1\\ 0 & 1 & -1\end{pmatrix}$.

Several solutions to estimate the angle $\theta$ are proposed in the literature. In [5] an off-line calibration step is proposed. It uses a color pattern or a set of preregistered images showing illumination changes. A 'self-calibration' approach is also presented. The strategy consists to find the angle for which the entropy of the invariant image is minimum. The later method is proved to be capable of finding the correct angle by using a single image where remarkable shaded areas are present. Despite its effectiveness in some cases, there are situations where it fails (*e.g.* in the presence of global illumination changes or soft shadows).

## 3. Combining AAMs and Light-Invariance

There are several possible ways of combining the AAMs with the Light-Invariant transformation $\mathcal{L}$. The general idea is to project the generated color image and the test color image to the Light-Invariant space, and to match them within that space:

$$\sum_x \left[A_0^{LI}(x) + \sum_{i=1}^{m}\lambda_i A_i^{LI}(x) - I^{LI}(W(x;p))\right]^2, \tag{9}$$

where $A_0^{LI}$, all $A_i^{LI}$ and $I^{LI}$ are Light-Invariant images, pre-transformed from the 3D color space to the 1D Light-Invariant space. We therefore get a classical 1D dataset, and the problem resembles the one in (5) for greylevel images.

The main drawback of this approach, which limits the system usability, is that we need to know in advance the angle for the test data. As we explained above, there exist calibration methods but their applicability is restricted in several practical cases. The best thing would be to set up an approach which does not require a specific off-line calibration step.

Following the same philosophy as in [10], a more general solution is proposed by embedding the angles within the cost function. Under this approach, the angles can be estimated together with the shape and appearance parameters of the AAM.

To do so, the training of appearance is done in the logRGB space described in §2. It is linearly mapped to the invariant space with (8) and, as described in §6, allows us to reconstruct the appearance in RGB.

The resulting cost function is:

$$\sum_x \Bigg[L(\theta_1)\left(\tilde{A}_0(x) + \sum_{i=1}^{m}\lambda_i\tilde{A}_i(x)\right) - L(\theta_2)(I(W(x;p))\Bigg]^2, \tag{10}$$

where $I$ represents the logRGB test image, $\tilde{A}_i$s are the logRGB appearance components, $\theta_1$ and $\theta_2$ are the angle parameters for the training sequence and the test image respectively. Using two different angles allows the system to work with different cameras (or the same camera with different photometric adjustments) for training and testing.

We examine two scenarios:

**Case 1: Photometric 'self-calibration' of the training camera.** The AAM can be 'self-calibrated' while fitting. If the training and test images are taken with the same camera $\theta_1 = \theta_2$. If two different cameras are used, we generally have $\theta_1 \neq \theta_2$. It is strictly necessary in both cases that changes in illumination arise between the training and the test images. If not, the angle or angles cannot be obtained and move randomly during optimization.

**Case 2: Basic Light-Invariant AAM (LI-AAM) fitting.** We assume that $\theta_1$ is known (the AAM is photometrically calibrated). In practice, $\theta_2$ is rarely known. We estimate it while fitting the AAM in the Light-Invariant space.

## 4. Basic Light-Invariant AAM fitting

The basic Light-Invariant AAM (LI-AAM) fitting includes the angle $\theta_2$, needed to convert the test images to the invariant space. As we stated before the training is performed in logRGB so a linear appearance basis is obtained within that space.

The appearance basis is made of an average vector $\tilde{A}_0$ and the set of $m$ appearance vectors $\tilde{A}_i$. The cost function is (10). It is minimized over $p$, $\lambda$ and $\theta_2$ using an additive Gauss-Newton algorithm. We refer the reader to [3] for more details.

## 5. Photometric Calibration of a Regular AAM

There are two possible situations:

- The same camera took the testing and training images. The cost function is minimized over $p$, $\lambda$ and $\theta_1 = \theta_2$
- Different cameras took the testing and training images. The cost function is minimized over $p$, $\lambda$, $\theta_1$ and $\theta_2$

## 6. Shadow-Free Appearance Reconstruction

Fitting the AAM in the Light-Invariant space not only allows to handle uncontrolled illuminations but it also permits to retrieve the face appearance under the canonical illumination (the one in the training set). We conveniently choose to transform the RGB data to the logRGB since it makes the transformation 'more linear'. The second advantage of our choice is that when the AAM correctly fits the face on the input image, the synthesized logRGB appearance can easily be back-converted to the original RGB space.

Assuming that the color components are strictly positive, the logRGB space is invertible to RGB. Given the set of appearance parameters $\lambda$ and the appearance basis in logRGB $\tilde{A}_i, \cdots, \tilde{A}_m$, the reconstructed RGB is:

$$A(x) = e^{\left(\tilde{A}_0(x) + \sum_{i=0}^{m} \lambda_i \tilde{A}_i(x)\right)} \quad (11)$$

where $e$ is the element-wise exponential.

Such an operation is usually ill-posed when no prior information is provided on the image content [5]. In our case the AAM gives a unique solution.

Shadow-free appearance reconstruction can be useful for face recognition systems usually, exhibiting better performances under controlled illumination, and for image compression applications, where it is useful to encode facial and illumination data separately.

## 7. Experimental Results

To evaluate the performance of the above presented algorithms, we tested their robustness to noise and their resistance to initial geometric shape displacement. We measured a geometric fitting error using manually labelled images and the angle error to the ground-truth. So as to allow the reader to compare with various existing methods, we study the results of RGB, greylevel, and different Light-Invariant (LI) AAM fitting procedures:

- **LI-AAM:** Basic Light Invariant AAM fitting, proposed in §4, where only the input image angle $\theta_2$ is unknown.
- **LI-AAM(LIT):** Basic Light Invariant AAM fitting, proposed in §4, where only the input image angle $\theta_2$ is unknown but performing the training in LI space.
- **LI-AAM(PC):** Photometric calibration of the AAM proposed in §5. Both $\theta_1$ and $\theta_2$ are obtained.
- **LI-AAM(PC):** Photometric calibration of the AAM proposed in §5, imposing the constraint $\theta_1 = \theta_2$.
- **RGB-AAM:** A regular AAM fitted in color space.
- **Grey-AAM:** A regular AAM fitted in greylevel space.

### 7.1. Simulated Perturbations

The whole set of algorithms are tested against image intensity noise and initial geometric displacement.

A person-specific AAM is trained using 6 face images taken under a mainly homogeneous illumination, and different poses.

The original images of the database are corrupted by noise with variance $\sigma_n$ over the range $[0 - 255]$ and on each color channel. One of the images used for training is corrupted by noise and an artificial global change of illumination is created. The shadow area modifies the RGB color values of the image to simulate the response of a camera with angle $\theta = 146$ degrees perturbed by a light source with a given color temperature $T$ (See Figure 2.a). The initial position of the model mesh is obtained by displacing its points from the hand labelled solution by a random distance, the variance of which denoted $\gamma$.


(a) (b)

Figure 2. (a) Artificial perturbation on an original RGB image. (b) Extract of the homogeneously illuminated face images used for training.

We compare the performance of all algorithms for a range of perturbation parameters $\gamma$ and $\sigma_n$. Figure 3.a shows the geometric error measured after a sufficient number of iterations to let each AAM converge (we used 50 iterations). On Figure 3.b and Figure 3.c, the geometric error and the angle error are tested against the noise (also measured after 50 iterations).

We observed that:

- LI-AAM and LI-AAM(LIT) perform exactly the same in all tests, so the use of logRGB for training appearance do as good as the direct use of the LI space.
- The image noise substantially affects all the proposed Light-Invariant methods. However LI-AAM shows to be the most robust.
- The initial geometric displacement substantially and equivalently affects all the Light-Invariant methods.

### 7.2. Real Illumination Changes

We tested the proposed solution on two sets of real data. Each set is composed of two sequences, both presenting one character changing head pose under neutral expression and finally smiles. On one sequence, the face is taken under homogeneous illumination as shown on Figure 2.b.

The presented color images are used to train the various AAMs. On the other sequence, the face is illuminated laterally as shown on Figure 4, and we compare the AAMs performance on this sequence. Figure 5 shows the evolution of the geometric error against the iteration number. As an observation, both LI-AAM and LI-AAM(LIT) present (equivalently) better convergence characteristics than the other LI methods. Also it clearly appears how the Grey-AAM and RGB-AAM quickly diverge.

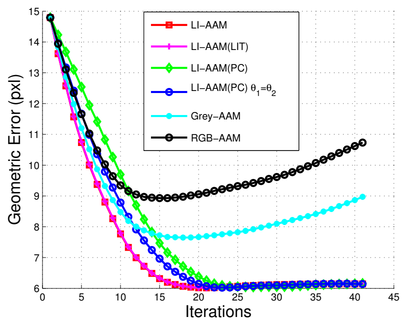



Figure 5. Geometric Error versus iteration number for the test image shown in Figure 4.a

## 8. Discussion

We tackled the difficult problem of Active Appearance Model fitting in an unconstrained illumination context. An AAM in its basic formulation is very sensitive to an unknown illumination.

Instead of classically matching the model appearance to the input image in the color space, we project the images, both from the synthesized and the input ones into a Light-Invariant space where the effect of shading are canceled. This only requires to tune a camera dependent angle $\theta$, for the cameras used for shooting the training and the test datasets. The calibration of these parameters, necessary to the Light-Invariant transformation, is done jointly with the appearance and shape parameters. This is what we called the Light-Invariant AAM fitting.

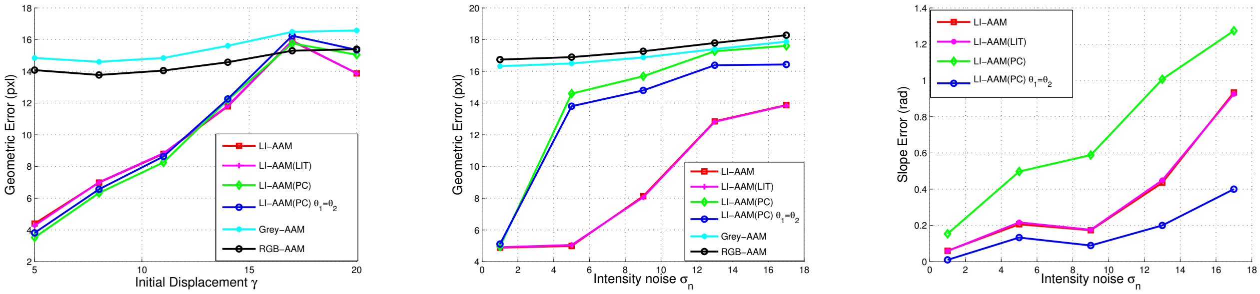
Two different scenarios are proposed: in the first acts the basic LI-AAM algorithm for which only the angle corresponding to the input image is unknown, and in the second is involved the photometric calibration of an existing AAM (LI-AAM(PC)) for which the angles corresponding to both the training and the test data (the cameras used to take them) are sought after.

Our approach does not need that the camera is precalibrated and makes the system useful to unsourced (color) image databases. In addition, we show how training the AAM in logRGB allows to reconstruct the RGB shadow-free appearance of a face after the fitting is performed.

Experimental results show that our proposal outperforms the regular AAM approaches (using RGB or greylevel values), when the input images present different illumination conditions.

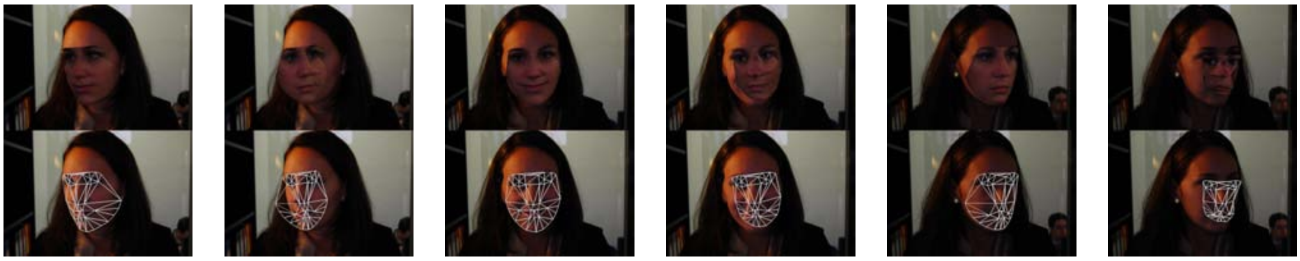
## References


[1] S. Baker, R. Gross, I. Matthews, and T. Ishikawa. Lucas-kanade 20 years on: A unifying framework: Part 2. Technical Report CMU-RI-TR-03-35, Robotics Institute. Carnegie Mellon University, Pittsburgh PA, 2003.

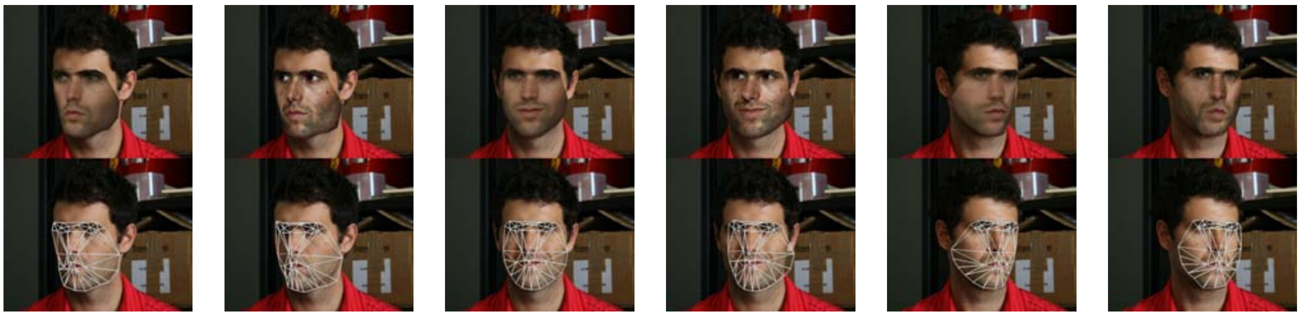



a) Geometric error versus $\gamma$ | b) Geometric error versus $\sigma_n$ | c) Angle error versus $\sigma_n$

Figure 3. Experiments using synthetic perturbations



a) LI-AAM | b) RGB | c) LI-AAM | d) RGB | e) LI-AAM | f) RGB



g) LI-AAM | h) RGB | i) LI-AAM | j) RGB | k) LI-AAM | m) RGB

Figure 4. Fitting on real data

[2] S. Baker and I. Matthews. Equivalence and efficiency of image alignment algorithms. *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, 1:1090–1097, 2001.

[3] S. Baker and I. Matthews. Lucas-kanade 20 years on: A unifying framework. *International Journal of Computer Vision*, 56(3):221–255, 2004.

[4] T. Cootes, G. Edwards, and C.J. Taylor. Active appearance models. *Transactions on Pattern Analysis and Machine Intelligence*, 23(6):681–685, june 2001.

[5] Graham D. Finlayson, Steven D. Hordley, Cheng Lu, and Mark S. Drew. On the removal of shadows from images. *Transactions on Pattern Analysis and Machine Intelligence*, 28:59 – 68, January 2006.

[6] F. Kahraman, M. Gokmen, S. Darkner, and R. Larsen. An active illumination and appearance (aia) model for face alignment. In *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, June 2007.

[7] S. Kim and M. Pollefeys. Robust radiometric calibration and vignetting correction. *Transactions on Pattern Analysis and Machine Intelligence*, 30(4):562–576, 2008.

[8] S.J. Kim and M. Pollefeys. Radiometric self-alignment of image sequences. In *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, 2004.

[9] I. Matthews and S. Baker. Active appearance models revisited. *International Journal on Computer Vision*, 60(2):135–164, November 2004.

[10] D. Pizarro and A. Bartoli. Shadow resistant direct image registration. In *Scandinavian Conference on Image Analysis*, Aalborg, Denmark, June 2007.

[11] T. Sim and T. Kanade. Combining models and exemplars for face recognition: An illuminating example. *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, 2001.

[12] S. Yan, C. Liu, S.Z. Li, H. Zhang, H.Y. Shum, and Q. Cheng. Face alignment using texture-constrained active shape models. *Image and Vision Computing*, 21(1):69–75, 2003.