Fourier Active Appearance Models

Rajitha Navarathna, Sridha Sridharan
Queensland University of Technology, Australia.
r.navarathna@qut.edu.au, s.sridharan@qut.edu.au

Simon Lucey
Commonwealth Scientific and Industrial Research Organisation (CSIRO), Australia.
simon.lucey@csiro.au

Abstract

Gaining invariance to camera and illumination variations has been a well investigated topic in Active Appearance Model (AAM) fitting literature. The major problem lies in the inability of the appearance parameters of the AAM to generalize to unseen conditions. An attractive approach for gaining invariance is to fit an AAM to a multiple filter response (e.g. Gabor) representation of the input image. Naively applying this concept with a traditional AAM is computationally prohibitive, especially as the number of filter responses increase. In this paper, we present a computationally efficient AAM fitting algorithm based on the Lucas-Kanade (LK) algorithm posed in the Fourier domain that affords invariance to both expression and illumination. We refer to this as a Fourier AAM (FAAM), and show that this method gives substantial improvement in person specific AAM fitting performance over traditional AAM fitting methods.

1. Introduction

Active Appearance Models (AAMs) [5, 21] employ a paradigm of inverting a synthesis model (or in machine learning terms a generative model) of how an object can vary in terms of shape and appearance. As a result, the ability of AAMs to register an unseen object image is intrinsically linked to how well the synthesis model can reconstruct the object image. Unfortunately, from a registration perspective, AAMs have inherent problems when attempting to fit to "real-world" objects which often have substantial shape and appearance variation.

Gross et al. [14] demonstrated that this problem is especially problematic in AAM face fitting. Specifically, Gross et al. showed that: (i) person specific AAMs substantially outperform a generic (i.e. models trained across many subjects) AAM, and (ii) this disparity in performance stems from the poor generalization properties of the appearance model of the generic AAM.

Generic non-rigid face fitting is still an ongoing topic in computer vision with notable theoretical inroads being made [8, 19, 23]. However, none of these approaches can provide the level of registration accuracy or computational efficiency achieveable through a person specific AAM [4, 14]. As a result, person specific AAMs are still the method of choice in a number of applications where users are willing to provide subject specific images and labels. Notable applications of person specific AAMs in literature can be found in areas such as expression classification, avatar synthesis, and visual speech synthesis [4].

The Problem: Even though state-of-the art person specific AAM face fitting outperforms generic non-rigid face fitting methods, significant problems still remain. A major drawback to person specific AAMs stems from their ability to only generalize to small amounts of appearance variation (essentially appearance variation that can be expressed as a linear combination of the training instances, e.g. expression variation). When unaccounted appearance variations are encountered due to a change in the environment (e.g., illumination or camera change), person specific AAMs perform poorly. This effect severely limits the usefulness of person specific AAMs, as one either needs to: (i) ensure the environment is strictly controlled, or (ii) collect and label training examples of the subject in the new environment.

Contributions: It has been well documented [21] that AAMs can be efficiently fitted through extensions to the classic Lucas & Kanade (LK) algorithm [20]. Of particular importance to LK inspired AAM fitting are the inverse compositional "simultaneous" and "project-out" extensions to the LK algorithm [21]. Recently, a new extension was proposed by Ashraf and Lucey [1] demonstrating how the traditional LK algorithm can be posed in the Fourier domain. This approach, which the authors refer to as Fourier

LK (FLK), is advantageous over traditional LK as doing image alignment on a high dimensional bank of filter response images is mathematically equivalent to doing alignment in the low dimensional raw image pixel space, if appropriate weightings are applied in the Fourier domain.

The main contributions of this paper are as follows:

- We show how LK inspired AAM fitting gives identical performance in the spatial and Fourier domains. Further, we demonstrate how the effect of multiple filter responses can be re-interpreted as a diagonal weighting matrix in the Fourier domain leading to substantial computational savings when performing inverse compositional simultaneous fitting across multiple filter responses (Section 5).
- We demonstrate the process of applying the robust error function in the Fourier domain by showing how: (i) the Fourier transform to the current image, and (ii) show the effect of multiple filter responses can be re-interpreted as a diagonal weighting matrix in the Fourier domain as simultaneous algorithm. (Sections 3 and 5)
- We empirically show the substantial improvement in person specific AAM fitting performance over canonical LK inspired fitting algorithms (i.e. (a) simultaneous and (b) simultaneous with a robust error function in Fourier domain), when using our proposed Fourier variants. For all our experiments we employed biologically motivated Gabor filter banks. (Sections 6 and 7)

Related work: Gaining invariance to environmental variations such as camera and illumination variations has been a well investigated topic in AAM fitting literature [7, 14, 24]. Notably, Gross et al. [14] modeled illumination variation by using an abundant of examples from different illumination conditions. As discussed earlier, this approach is unattractive in practice as one has to collect multiple images/labels of the subject from a wider variety environmental conditions. Recently, Theobald et al. [24] demonstrated the usefulness of robust-error functions for AAM fitting for dealing with previously unseen appearance variations. Although successful, this approach is problematic as it requires a recomputation of the Hessian for each iteration of fitting irrespective of the approach employed (i.e., simultaneous and project-out).

Filter-based solutions have also been utilized in the past to gain environmental invariance in AAM fitting. Of particular note is the work of Cootes and Taylor [7] where the authors explored the use of multiple filter (specifically orientated gradients) responses for fitting. Although exhibiting impressive results, the approach is problematic as it requires the explicit computation of multiple image filter responses at each iteration of AAM fitting. Our work differs to the work presented in [7] in that we are proposing a novel method for completely pre-computing the effect of multiple filter responses such that the online portion of the AAM fitting algorithm operates solely and efficiently on raw pixels.

General Notation: Vectors are always represented in lower-case bold (e.g., $\mathbf{a}$). Matrices are always expressed in upper-case bold (e.g., $\mathbf{A}$). Scalars in lower-case (e.g. a). Images in this paper shall always be expressed in capitalized form A. Warp functions $\mathcal{W}(\mathbf{x};\mathbf{p})$ will be used throughout this paper to denote a warping of a $2D$ coordinate vector $\mathbf{x} = [x, y]^T$ by a warp parameter vector $\mathbf{p} \in \mathcal{R}^P$, where P is the number of warp parameters, back to a fixed base coordinate system. This base coordinate system is defined when $\mathbf{p} = \mathbf{0}$ such that $\mathcal{W}(\mathbf{x};\mathbf{p}) = \mathbf{x}$. An abuse of notation is entertained in this paper for when an image A is warped by the warp parameter vector $\mathbf{p}$, such that $A(\mathbf{p}) = [A(\mathcal{W}(\mathbf{x}_1;\mathbf{p})), \ldots, A(\mathcal{W}(\mathbf{x}_D;\mathbf{p}))]^T$. In this instance $A(\mathbf{p})$ is a D dimensional vector of image intensities, where D denotes the number of discrete coordinates in the base coordinate system. The steepest descent matrix $\frac{\partial A(\mathbf{p})}{\partial \mathbf{p}}$ of an image $A(\mathbf{p})$ is used frequently through out this paper. This $D \times P$ matrix is formed by combining image gradients of $A(\mathbf{p})$ with the Jacobian of the warp function $\mathcal{W}(\mathbf{x};\mathbf{p})$, more details on the formation of this matrix can be found in [21]. Finally, we use the notation $\| \mathbf{a} \|^2_{\mathbf{Q}}$ to represent the quadratic form $\mathbf{a}^T\mathbf{Q}\mathbf{a}$, and $\mathbf{Q}$ is a symmetric, positive semi-definite weighting matrix.

Fourier Notation: This paper also borrows heavily upon concepts from signal processing. A 2D convolution operation is represented as the $*$ operator. A $\hat{}$ applied to any vector denotes the 2D Discrete Fourier Transform (DFT) of a vectorized 2D image $A(\mathbf{p})$ or signal $\mathbf{a}$ such that $\hat{A}(\mathbf{p}) \leftarrow \mathbf{F}\, A(\mathbf{p})$ and $\hat{\mathbf{a}} \leftarrow \mathbf{F}\mathbf{a}$. $\mathbf{F}$ is the $D \times D$ matrix of complex basis vectors for mapping to the Fourier domain for any D dimensional vectorized image/signal. We have chosen to employ a Fourier representation in this paper due to its particularly useful ability to represent convolutions as a Hadamard product in the Fourier domain. Additionally, we take advantage of the fact that $\text{diag}(\hat{\mathbf{g}})\hat{\mathbf{a}} = \hat{\mathbf{g}} \circ \hat{\mathbf{a}}$, where $\circ$ represents the Hadamard product, and $\text{diag}()$ is an operator that transforms a D dimensional vector into a $D \times D$ dimensional diagonal matrix. The role of filter $\hat{\mathbf{g}}$ or signal $\hat{\mathbf{a}}$ can be interchanged with this property. Any transpose operator ^{T} on a complex vector or matrix in this paper additionally takes the complex conjugate in a similar fashion to the Hermitian adjoint [22].

2. Active Appearance Models

Active appearance models (AAMs) [6, 21] are usually constructed from a set of training images with the AAM mesh vertices hand-labeled on them [6]. The training mesh vertices are first aligned with procrustes analysis. Then principal component analysis (PCA) is used to build a 2D

linear model of shape variation [6]. The shape $\mathbf{s}$ of an AAM is described by a 2D triangulated mesh. The 2D shape $\mathbf{s} = (x_1, y_1, \ldots, x_v, y_v)^T$ can be represented as a base shape $\mathbf{s}_0$ plus a linear combination of P shape vectors $\mathbf{s}_i$:

$$\mathbf{s} = \mathbf{s}_0 + \sum_{i=1}^{P} p_i \mathbf{s}_i \tag{1}$$

where $\mathbf{p} = [p_1, \ldots, p_P]^T$ is the shape parameter vector. The AAM model of appearance variation is obtained by first warping all the training images onto the mean shape and then applying PCA on the shape normalized appearance images. The appearance of an AAM $A(\mathbf{0})$ is an image vector defined over the pixels $\mathbf{x} \in \mathbf{s}_0$ inside the base mesh $\mathbf{s}_0$ when $\mathbf{p} = \mathbf{0}$. The appearance $A_{\boldsymbol{\lambda}}(\mathbf{0})$ can be represented as a mean appearance $A_0(\mathbf{0})$ plus a linear combination of K orthonormal appearance vectors $A_j(\mathbf{0})$:

$$\begin{aligned} A_{\boldsymbol{\lambda}}(\mathbf{0}) &= A_0(\mathbf{0}) + \sum_{j=1}^{K} \lambda_j A_j(\mathbf{0}) \\ &= A_0(\mathbf{0}) + \mathbf{A}\boldsymbol{\lambda} \end{aligned} \tag{2}$$

where $\boldsymbol{\lambda} = [\lambda_1, \ldots, \lambda_K]^T$ is the appearance parameter vector and $\mathbf{A} = [A_1(\mathbf{0}), \ldots, A_K(\mathbf{0})]$ is the matrix of concatenated appearance vectors.

3. LK inspired AAM fitting

A number of approaches have been proposed in literature for fitting AAMs [6, 21]. The most notable and popular of these variants are approaches based on the Lucas & Kanade (LK) algorithm [21]. In this approach one can pose AAM fitting as minimizing the following objective function:

$$\arg\min_{\mathbf{p},\boldsymbol{\lambda}} \parallel I(\mathbf{p}) - A_0(\mathbf{0}) - \mathbf{A}\boldsymbol{\lambda} \parallel^2_{\mathbf{Q}} \tag{3}$$

where $I(\mathbf{p})$ represents the warped input image using the warp specified by the parameters $\mathbf{p}$.

The central task of the objective function described in Equation 3 is to find the shape $\mathbf{p}$ and appearance $\boldsymbol{\lambda}$ that minimizes the weighted sum of squared distances (SSD) between the warped input image and the AAM. For most AAM fitting problems the weight matrix $\mathbf{Q}$ is assumed to be an identity matrix $\mathbf{I}$ (i.e. unweighted SSD).

Generally, the objective function in Equation 3 is difficult to solve as their is a non-linear relationship between the shape $\mathbf{p}$, and appearance $\boldsymbol{\lambda}$ parameters. A key insight, stemming from Lucas & Kanade [20], was that a linear approximation can be made between $\mathbf{p}$ and $\boldsymbol{\lambda}$ through the judicious use of image gradients and the chain rule to form steepest descent matrices (i.e. $\frac{\partial A(\mathbf{p})}{\partial \mathbf{p}}$). In this section we will briefly review two common approaches in AAM fitting.

Simultaneous algorithm: The simultaneous algorithm [21] linearizes the objective function in Equation 3 such that:

$$\arg\min_{\Delta\mathbf{p},\Delta\boldsymbol{\lambda}} \parallel I(\mathbf{p}) - A_{\boldsymbol{\lambda}}(\mathbf{0}) - \frac{\partial A_{\boldsymbol{\lambda}}(\mathbf{0})}{\partial \mathbf{p}}\Delta\mathbf{p} - \mathbf{A}\Delta\boldsymbol{\lambda} \parallel^2_{\mathbf{Q}} \quad . \tag{4}$$

Instead of solving for the shape $\mathbf{p}$ and appearance $\boldsymbol{\lambda}$ parameters directly, through the linearization step in 4 we iteratively solve for the updates $\Delta\mathbf{p}$ and $\Delta\boldsymbol{\lambda}$. The objective function in Equation 4 takes advantage of a computationally efficient extension to the LK algorithm referred to as the inverse compositional (IC) algorithm [21]. The IC algorithm differs from the canonical LK algorithm as it linearizes the template image $A_{\boldsymbol{\lambda}}(\Delta\mathbf{p})$, with respect to $\Delta\mathbf{p}$, instead of the source image $I(\mathbf{p} + \Delta\mathbf{p})$.

A consequence for this switch is that the update to the the current warp parameters are updated by the inverse (as we want to update the source image not the template) of the warp update $\mathbf{p} \leftarrow \mathbf{p} \circledcirc \Delta\mathbf{p}^{-1}$. The operation $\circledcirc$ represents the composition of two warps (e.g. for an affine warp this is represented as a matrix multiplication). The update to the appearance parameters, however, remain additive such that $\boldsymbol{\lambda} \leftarrow \boldsymbol{\lambda} + \Delta\boldsymbol{\lambda}$. The explicit solution to $\Delta\mathbf{p}$ and $\Delta\boldsymbol{\lambda}$ can be found "simultaneously" such that:

$$\begin{bmatrix} \Delta\mathbf{p} \\ \Delta\boldsymbol{\lambda} \end{bmatrix} = \mathbf{H}^{-1}_{sim}\mathbf{J}^T_{sim}\mathbf{Q}[I(\mathbf{p}) - A_{\boldsymbol{\lambda}}(\mathbf{0})] \tag{5}$$

where the pseudo simultaneous Hessian matrix is defined as $\mathbf{H}_{sim} = \mathbf{J}^T_{sim}\mathbf{Q}\mathbf{J}_{sim}$. The simultaneous Jacobian matrix is defined as:

$$\mathbf{J}_{sim} = \begin{bmatrix} \frac{\partial A_{\boldsymbol{\lambda}}(\mathbf{0})}{\partial \mathbf{p}} \\ \mathbf{A}^T \end{bmatrix} \quad . \tag{6}$$

Empirically, the simultaneous algorithm has been noted to have excellent fitting performance compared other LK inspired methods to AAM fitting. A major problem, however, with the simultaneous algorithm occurs with respect to computational efficiency. Specifically, as a consequence of the update step $\boldsymbol{\lambda} \leftarrow \boldsymbol{\lambda} + \Delta\boldsymbol{\lambda}$ the appearance image $A_{\boldsymbol{\lambda}}(\mathbf{0})$, Jacobian matrix $\mathbf{J}_{sim}$, and Hessian matrix $\mathbf{H}_{sim}$ must be re-estimated at each iteration.

Robust error function:

It is well known in that SSD cost criteria have problems in the presence of outliers (i.e pixels with the large reconstruction error). Recently, Theobald et al. [24] demonstrated the AAM fitting with a robust error functions for dealing with unseen appearance variations. The role of this method is to down weight the pixel outliers and minimise the Equation 3 using a robust error function with respect to the shape and the appearance parameters such that:

$$\arg\min_{\mathbf{p},\boldsymbol{\lambda}} \rho(\parallel I(\mathbf{p}) - A_{\boldsymbol{\lambda}}(\mathbf{0}) \parallel^2_{\mathbf{Q}}) \tag{7}$$

where $\rho(\cdot)$ is a robust error function. The $\Delta\mathbf{p}$ and $\mathbf{H}_{sim}$ have to be weighted by the error function $\rho'(\mathbf{e})$ or a weighting parameter w_i, where $\mathbf{e}$,

$$\mathbf{e} = (I(\mathbf{p}) - A_{\boldsymbol{\lambda}}(\mathbf{0}))^2 \tag{8}$$

This approach is problematic as it requires a recomputation of the Hessian for each iteration of fitting due to the changing in weighting parameter.

Weighted PCA: The appearance basis $\mathbf{A}$ is traditionally found using unweighted principal component analysis (PCA) to find the first K eigenvectors from raw pixel shape normalized training images. However, for the case when $\mathbf{Q} \neq \mathbf{I}$ the weighting matrix must be included in the canonical PCA objective function:

$$\arg\max_{\mathbf{A}} \text{tr}(\mathbf{A}^T\mathbf{V}\mathbf{C}\mathbf{V}^T\mathbf{A}) \text{ subject to } \mathbf{A}^T\mathbf{A} = \mathbf{I} \tag{9}$$

where $\mathbf{C}$ is the scatter matrix of the training images and $\mathbf{V}$ is the decomposition of the positive semi-definite weighting matrix $\mathbf{Q} = \mathbf{V}\mathbf{V}^T$.

4. AAM fitting on filter responses

Linear filters are often used to extract useful feature representations in computer vision. One particular filter, based on the seminal work of Gabor [13], that has received much attention in the vision community are Gabor wavelets due to their biological relevance and computational properties [10, 9, 11, 12]. The employment of a concatenation of Gabor filter responses, as a pre-processing step to deal with illumination change, before learning a classifier has found particular success in face identity [25, 18] and expression [3] recognition when compared to learning those classifiers with original appearance features/pixels.

Fitting an AAM across multiple filter linear filter responses involves minimizing the following objective function,

$$\arg\min_{\mathbf{p},\boldsymbol{\lambda}} \parallel \{\mathbf{g}_i * I(\mathbf{p})\}_{i=1}^{M} - \{\mathbf{g}_i * \mathbf{A}_{\boldsymbol{\lambda}}(\mathbf{0})\}_{i=1}^{M} \parallel^2 \tag{10}$$

where $\mathbf{g}_i$ is i-th filter with M filters in total, while $\{.\}_{i=1}^{M}$ represents the concatenation operation i.e. $\{\mathbf{x}_i\}_{i=1}^{M} = [\mathbf{x}_1^T \dots \mathbf{x}_M^T]^T$. One should note here that the weighting matrix $\mathbf{Q}$ has been omitted here, such that a $\mathbf{Q} = \mathbf{I}$ is assumed. The role of $\mathbf{Q}$ with respect to fitting across multiple filter responses shall be examined in Section 5.

Computational concerns: As pointed out by [2, 3, 18] a particular problem with Equation 10 is the inherently large memory and computational overheads required for representing images in this over-complete Gabor domain. Applying this strategy to the LK framework presents two fundamental problems. First, if there are M filters in the bank, and D pixels in the input image, we need to do M 2D convolutions involving images containing D pixels each. Second, the number of columns in the Jacobian $\mathbf{J}$ matrix for the simultaneous algorithm increases from D to MD. As a result of these computational overheads, the idea of doing LK alignment with even a modest number of Gabor filter banks (e.g., 9 scales times 8 orientations, i.e. $M = 72$, as employed in [18]) becomes prohibitively expensive and impractical.

Even for smaller filter bank sizes authors in literature have resorted to methods for approximating the full response vectors such as: (i) downsampling of filter responses [18], (ii) employing filter responses at certain fiducial positions within the image [25], (iii) the employment of feature selection methods to select the most discriminative filter responses [3], and most recently (iv) where individual classifiers are learnt for each filter response and a fusion strategy employed to combine the outputs in a synergistic manner [17].

5. FLK inspired AAM fitting

Recently, Ashraf and Lucey [1] proposed an extension to the LK algorithm for fitting a template across multiple filter responses that circumvents most of these computational concerns. In this section we have extended this work specifically to the case of AAM fitting.

It is elementary to show that the error in Equation 10 can equivalently be written as:

$$\arg\min_{\mathbf{p},\boldsymbol{\lambda}} \sum_{i=1}^{M} \parallel \mathbf{g}_i * [I(\mathbf{p}) - A_{\boldsymbol{\lambda}}(\mathbf{0})] \parallel^2 \quad . \tag{11}$$

Exploiting the fact that convolution becomes a Hadamard (i.e., element-by-element) product in the Fourier domain, and employing Parseval's relation [22] (energy content is preserved as we move from the spatial to the Fourier domain), we may write the error in Equation 11 as follows:

$$\arg\min_{\mathbf{p},\boldsymbol{\lambda}} \parallel \hat{I}(\mathbf{p}) - \hat{A}_{\boldsymbol{\lambda}}(\mathbf{0}) \parallel^2_{\mathbf{S}} \tag{12}$$

where,

$$\mathbf{S} = \sum_{i=1}^{M} (\text{diag}(\hat{\mathbf{g}}_i))^T \text{diag}(\hat{\mathbf{g}}_i) \tag{13}$$

and $\hat{I}(\mathbf{p}), \hat{A}_{\boldsymbol{\lambda}}(\mathbf{0}), \hat{\mathbf{g}}_i$ are the $2D$ Fourier transforms of vectorized images $I(\mathbf{p}), A_{\boldsymbol{\lambda}}(\mathbf{0})$ and filters $\mathbf{g}_i$ respectively. The matrix $\mathbf{S}$ is a *diagonal* matrix that can be *precomputed* and is *independent* of the number of filters being applied. We also know that the operation of a $2D$ Fourier transform can be replaced by pre-multiplying a signal (of length D) by a $D \times D$ matrix $\mathbf{F}$ containing the Fourier basis vectors. This can be seen in the following FLK objective function,

$$\arg\min_{\mathbf{p},\boldsymbol{\lambda}} \parallel I(\mathbf{p}) - \mathbf{A}_{\boldsymbol{\lambda}}(\mathbf{0}) \parallel^2_{\mathbf{F}^T\mathbf{S}\mathbf{F}} \quad . \tag{14}$$

Fourier Simultaneous : An immediate consequence of Equation 14 is that it now becomes possible to apply the canonical simultaneous and simultaneous with a robust error function fitting algorithms, described in Section 4, by setting the weight matrix to:

$$\mathbf{Q} = \mathbf{F}^T\mathbf{S}\mathbf{F} \tag{15}$$

where $\mathbf{S}$ (Equation 13) is determined by the choice of filters being used. Moreover we can also see that FLK and LK inspired fitting strategies become equivalent when $\mathbf{S} = \mathbf{I}$ since $\mathbf{F}^T\mathbf{F} = \mathbf{I}$[1].

FAAM with a Fourier robust error function:

We can write the Equation 12 with a robust error function with the $2D$ Fourier transforms of vectorized images $I(\mathbf{p}), A_{\boldsymbol{\lambda}}(\mathbf{0})$ in Fourier domain.

$$\arg\min_{\mathbf{p},\boldsymbol{\lambda}} \rho(\| \hat{I}(\mathbf{p}) - \hat{A}_{\boldsymbol{\lambda}}(\mathbf{0}) \|^2_{\mathbf{S}}) \tag{16}$$

The weighting parameter is given by $\rho'(\hat{\mathbf{e}})$. It has been shown for AAM fitting [24] that use of a exponential function $\rho'(\mathbf{e}) = exp(-c.\mathbf{e})$ as the robust error function gives better performance in Euclidean AAM, given that c is suitably selected. We re-formulate this with a Fourier exponential function (as robust error function) such that $\rho'(\hat{\mathbf{e}}) = exp(-c.\hat{\mathbf{e}})$ and the weights for the each iteration are estimated using,

$$w_i \leftarrow w_i \exp(-c.\hat{\mathbf{e}}) \tag{17}$$

Computational concerns:

By casting the AAM algorithm in the Fourier domain, we have shown that it is equivalent to the AAM with a weighting matrix $\mathbf{Q} = \mathbf{F}^T\mathbf{S}\mathbf{F}$. In practice, however, one never explicitly computes $\mathbf{Q}$, instead applying efficient DFTs to the source and appearance images directly. For the simultaneous algorithm, this has the small drawback of having to perform a DFT at each iteration of the algorithm adding to its already sizable computational cost. However, what makes this approach computationally feasible is that we can replace the matrix form of the Fourier transform $\mathbf{F}$ which has a cost of $O(N^2)$ with a computationally feasible Fourier transform which is $O(NlogN)$ [22], where N is the number of pixels.

Computational cost of most of the steps depends on (i) n number of warp parameters and (ii) m number of appearance parameters. The computational cost is independent from the number of Gabor filters. Table 1 shows the summary of the computational cost.

[1]It should be noted that in many practical formulations of a 2D-DFT $\mathbf{F}^T\mathbf{F} = c\mathbf{I}$, where c is a constant. Typically, $c = D$ where D is the dimensionality of the feature space. This detail has been omitted in the main portion of this paper for the sake of clarity.

Step	Complexity
Warp I with $\mathbf{p}$ to compute $I(\mathbf{p})$	$O(nN)$
Compute the error image: $I(\mathbf{p}) - \mathbf{A}_{\boldsymbol{\lambda}}(\mathbf{0})$	$O(mN)$
Compute FFT of the error image	$O(NlogN)$
Compute the steepest descent images	$O((n+m)N)$
Compute the Jacobian	$O((n+m)N)$
Compute FFT for the Jacobian	$O((n+m)NlogN)$
Compute the Hessian $\mathbf{H}_{sim}$	$O((n+m)^2N)$
Compute the inverse of the Hessian	$O((n+m)^3N)$
Compute $\Delta\mathbf{q}$	$O((n+m)^2)$
Update $\mathbf{p} \leftarrow \mathbf{p} \circledcirc \Delta\mathbf{p}^{-1}$	$O(n^2)$
Update $\lambda \leftarrow \lambda + \Delta\lambda$	$O(m)$

Table 1. The computation cost of the Gabor FAAM algorithm.

6. MultiPIE Experiments

In order to compare the algorithms in terms of robustness to various illumination conditions, person specific AAM fitting experiments were conducted on the frontal subset of MultiPIE face database [15]. This consisted of 19 illumination conditions (i.e., 18 variations of flash firing and without flash). The database also consisted a range of facial expression including *neutral, smiles, surprise, squints, disgust and screams*. All images were hand annotated with 68 points. More details about the MulitiPIE database can be found in [15].

Throughout this section we will be comparing AAM fitting algorithms for two different weighting matrices: (i) $\mathbf{Q} = \mathbf{I}$, and (ii) $\mathbf{Q} = \mathbf{F}^T\mathbf{S}\mathbf{F}$ where $\mathbf{S}$ is defined through a bank of Gabor filters (9 scales times 8 orientations, see [1] for more details). We shall refer to all variants of (i) and (ii) as *Euclidean Active Appearance Models (AAM)* and *Gabor Fourier Active Appearance Models (FAAM)*.

Measuring fitting performance: For all our experiments, a person specific AAM was estimated for each subject in MultiPIE for frontal illumination. Two types of fitting performance were measured: (a) matched and (b) mismatched illumination. We measured fitting performance in terms of root mean square error (RMS) between the 2D mesh location of the current fit results and the ground-truth 2D mesh coordinates with respect to the base mesh. Results were calculated for (a) and (b) when the initialized shape was randomly perturbed from ground-truth.

Simultaneous results:

Figure 1 depicts the average RMS mesh location error against iterations for simultaneous variants of Euclidean AAMs and Gabor FAAMs for (a) matched and (b) mismatch illumination. Similarly, Figure 2 depicts the number of converged trials as a function of the RMS error threshold for (a) and (b). For (a) Euclidean AAM and Gabor FAAMs obtain almost identical performance. However, for (b) in the presence of mismatched illumination there is a clear advantage in using a Gabor FAAM.

FAAM with a robust error function results:

For these experiments, we reformulate the Equation 11

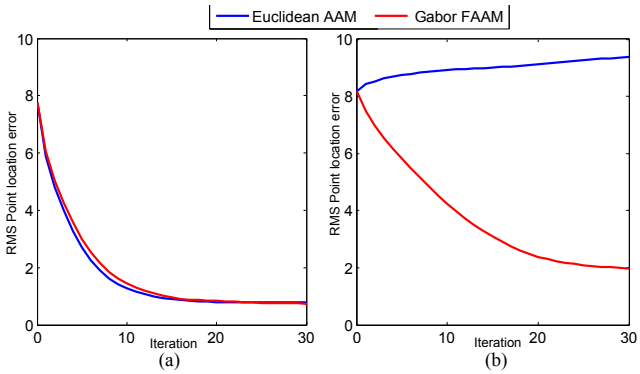


Figure 1. Average convergence rates for simultaneous algorithm: (a) when the input and training images have the same illumination conditions, both algorithms perform equally well. (b) when the illumination of the input image changes, the Gabor FAAM algorithm is still able to do the fitting, while the Euclidean AAM diverge.

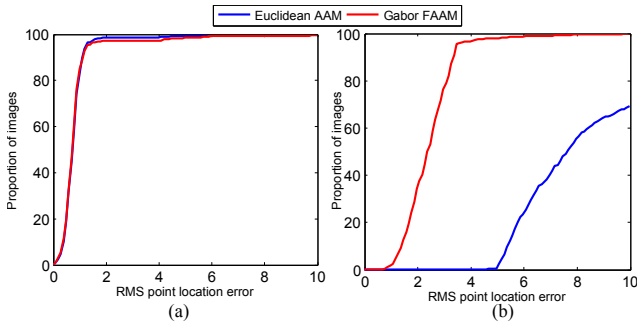


Figure 2. Fitting performance curves for simultaneous algorithm using Euclidean AAM and Gabor FAAM: (a) when the input and training images have the same illumination, (b) when the input and training images have the mismatched illumination.

with a robust error function such that $\rho'(\hat{\mathbf{e}}) = exp(-c.\hat{\mathbf{e}})$. Through a cross validation method we selected c as $c = 0.042$ for these experiments.

Figure 3 depicts the average RMS mesh location error against iterations for robust-error function variants of Gabor FAAM and Gabor FAAM with a robust error function for (a) matched and (b) mismatch illumination. Figure 4 depicts the number of converged trials as a function of the RMS error threshold for (a) and (b). In a similar fashion to the simultaneous results, (a) obtains almost identical performance in Gabor FAAM and Gabor FAAM with a robust error function. In the presence of substantial illumination mismatch (b) still both algorithms perform almost identical as depicted in Figures 3 and 4.

7. Tracking Experiments

We conducted various tracking experiments on video sequences containing substantial variations in illumination over time. An example of tracking sequence can be seen in

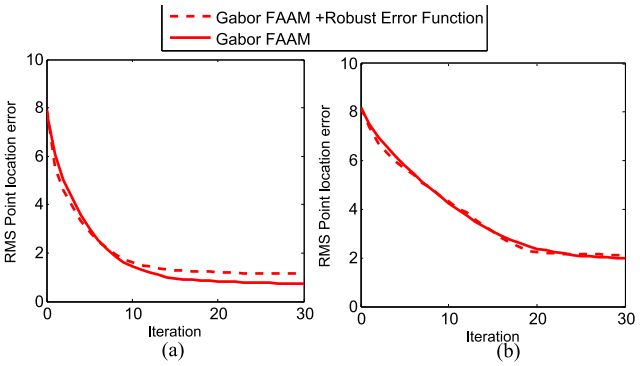


Figure 3. Average convergence rates for Gabor FAAM and Gabor FAAM with a robust error function: Both algorithms perform almost identical for (a) match illumination conditions (b) mis-match illumination condition.

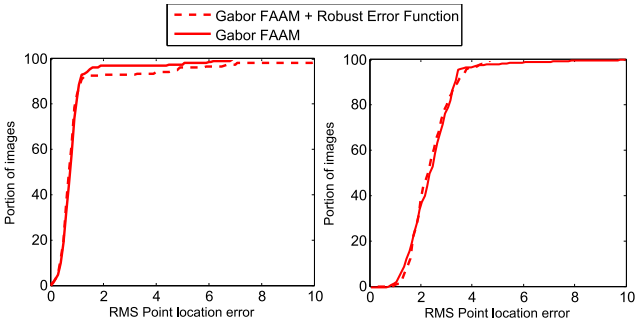


Figure 4. Fitting performance curves for Gabor FAAM and Gabor FAAM with a robust error function : (a) when the input and training images have the same illumination, (b) when the input and training images have the mismatched illumination.

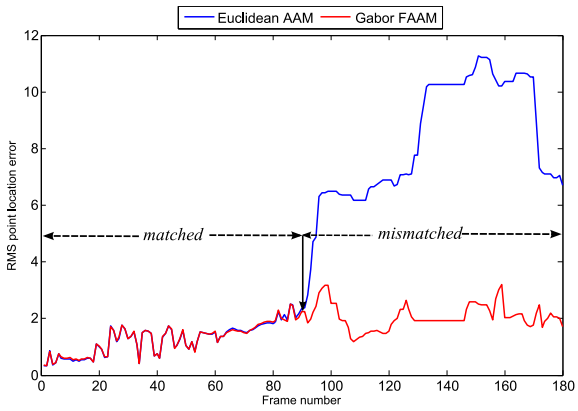
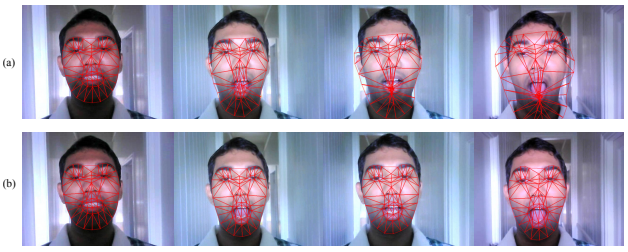


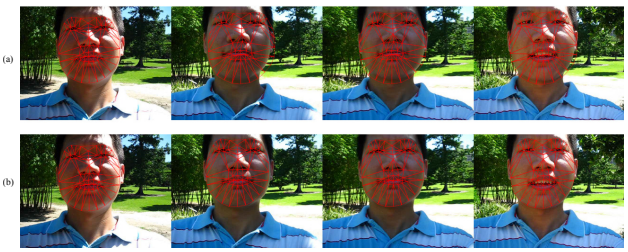
Figure 5. Example of a tracking with the Euclidean AAM and the Gabor FAAM in a video sequence. Illumination is changing over the time using the 3 different flashes. Euclidean AAM and the Gabor FAAM showed smiler results in the initial frames, but only Gabor FAAM showed good tracking results when the illumination changing over the time.

Figure 5. The sequence was obtained in a laboratory setting. Ground-truth for the first-frame was given for both the Euclidean AAM and Gabor FAAM. Results in terms of RMS

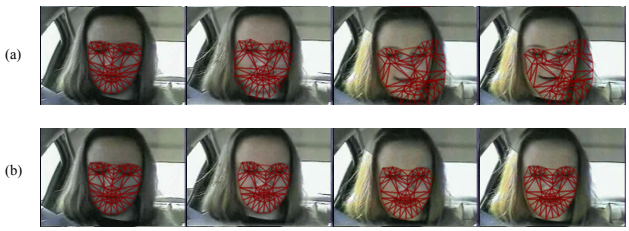


(a) Key frames taken from a video sequence of a person who is walking along a passage in a house.

(b) Key frames taken from a video sequence of a person who is walking along a passage in a building.



(c) Key frames taken from a video sequence of a person who is walking in a park.



(d) Tracking with sequence of image frames in a a real-world automobile environment for frames {1, 200, 350, 400}. Note: The video sequence was obtained from the AVICAR [16] database.

Figure 6. Challenging examples of tracking in a real world applications. Gabor FAAM showed good tracking results when the illumination changing over the time. Top Row : tracking sequence with the Euclidean AAM, Bottom row: tracking sequence with the Gabor FAAM

error from ground-truth can be seen in Figure 5 showing a substantial benefit to Gabor FAAM in person specific face tracking tasks. Visual examples of tracking performance in challenging environments can be seen in Figure 6.

8. Conclusion

In this paper we presented a novel extension to AAM fitting which we refer to as Fourier AAM. Some of our key contributions include: (i) demonstrating how LK inspired fitting gives identical fitting performance in the spatial and Fourier domains, (ii) show how the inverse compositional simultaneous algorithm can be posed in the Fourier domain with a robust error function, and (iii) show how Gabor FAAM gives dramatically improved performance over traditional AAM in the presence of unseen illumination variation.

Acknowledgment

This work was supported through the Cooperative Research Centre for Advanced Automotive Technology (AutoCRC) in Australia.

References

[1] A. B. Ashraf, S. Lucey, and T. Chen. Fast image alignment in the fourier domain. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2010*, pages 2480 –2487, June 2010. 1, 4, 5

[2] M. Bartlett, G. Littlewort, C. Lainscsek, I. Fasel, M. Frank, and J. Movellan. Fully automatic facial action recognition in spontaneous behavior. *7th International Conference on Automatic Face and Gesture Recognition*, pages 223–230, 2006. 4

[3] M. S. Bartlett, G. Littlewort, M. Frank, C. Lainscesk, I. Fasel, and J. Movellan. Recognizing facial expression: Machine learning and application to spontaneous behavior. *IEEE Conference on Computer Vision and Pattern Recognition*, 2:568–573, June 2005. 4

[4] J. F. Cohn. Advances in behavioral science using automated facial image analysis and synthesis. *IEEE Signal Processing Magazine*, 27(6), November 2010. 1

[5] T. Cootes, G. Edwards, and C. Taylor. Active appearance models. *In Proceedings European Conference on Computer Vision*, 2:484–498, 1998. 1

[6] T. Cootes, G. Edwards, and C. Taylor. Active appearance models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23, no. 6:681–685, 2001. 2, 3

[7] T. F. Cootes and C. J. Taylor. On representing edge structure for model matching. In *IEEE International Conference on*

Computer Vision and Pattern Recognition, volume 1, pages 1114–1119, 2001. 2

[8] D. Cristinacce and T. F. Cootes. Feature detection and tracking with constrained local models. In *British Machine Vision Conference (BMVC)*, pages 929–938, 2006. 1

[9] J. Daugman. Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional cortical filters. *Journal of the Optical Society of America*, 2(7):1160–1169, 1985. 4

[10] J. G. Daugman. Two-dimensional spectral analysis of cortical receptive field profiles. *Vision Research*, 20(10):847–856, 1980. 4

[11] J. G. Daugman. Complete discrete 2-D Gabor transforms by neural networks for image analysis and compression. *IEEE Trans. PAMI*, 36:1169–1179, July 1988. 4

[12] D. J. Field. Relations between the statistics of natural images and the response properties of cortical cells. *Journal of the Optical Society of America A*, 4(12):2379–2393, 1987. 4

[13] D. Gabor. Theory of communication. *Journal of the Institution of Electrical Engineers (London)*, 93(III):429–457, 1946. 4

[14] R. Gross, I. Matthews, and S. Baker. Generic vs. person specific active appearance models. *Image and Vision Computing*, 23(1):1080–1093, November 2005. 1, 2

[15] R. Gross, J. S. Baker, I. Matthews, and T. Kanade. Multi-PIE. In *IEEE International Conference on Automatic Face and Gesture Recognition*, 2008. 5

[16] B. Lee, M. Hasegawa-Johnson, C. Goudeseune, S. Kamdar, S. Borys, M. Liu, and T. Huang. AVICAR: An audiovisual speech corpus in a car environment. *In Proc. Interspeech 2004*, pages 2489–2492, Jeju Island, Korea. 7

[17] Z. Li, D. Lin, and X. Tang. Nonparametric discriminant analysis for face recognition. *IEEE Trans. PAMI*, 31(4):755–761, April 2009. 4

[18] C. Liu and H. Wechsler. Gabor feature based classification using the enhanced Fisher linear discriminant model for face recognition. *IEEE Trans. Image Processing*, 11(4):467–476, 2002. 4

[19] X. Liu. Generic face alignment using boosted appearance model. *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, 2007. 1

[20] B. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *International Joint Conference on Artificial Intelligence*, pages 674 – 679, 1981. 1, 3

[21] I. Matthews and S. Baker. Active appearance models revisited. *International Journal of Computer Vision*, 60(1):135 – 164, November 2004. 1, 2, 3

[22] A. V. Oppenheim and A. S. Willsky. *Signals & Systems*. Prentice Hall, 2nd edition, 1996. 2, 4, 5

[23] J. Saragih, S. Lucey, and J. F. Cohn. Deformable model fitting with a mixture of local experts. In *IEEE International Conference on Computer Vision (ICCV)*, pages 2248 – 2255, 2009. 1

[24] B. Theobald. Evaluating error functions for robust active appearance models. In *IEEE International Conference on Automatic Face and Gesture Recognition*, pages 149–154, 2006. 2, 3, 5

[25] C. Wiskott, J. M. Fellous, N. Krüger, and C. von der Malsburg. Face recognition by elastic bunch graph matching. *IEEE Trans. PAMI*, 19(7):775–779, July 1997. 4