

# Fourier Lucas-Kanade Algorithm

Simon Lucey, Rajitha Navarathna, Ahmed Bilal Ashraf, and Sridha Sridharan

**Abstract**—In this paper we propose a framework for both gradient descent image and object alignment in the Fourier domain. Our method centers upon the classical Lucas & Kanade (LK) algorithm where we represent the source and template/model in the complex 2D Fourier domain rather than in the spatial 2D domain. We refer to our approach as the Fourier LK (FLK) algorithm. The FLK formulation is advantageous when one pre-processes the source image and template/model with a bank of filters (e.g. oriented edges, Gabor, etc.) as: (i) it can handle substantial illumination variations, (ii) the inefficient pre-processing filter bank step can be subsumed within the FLK algorithm as a sparse diagonal weighting matrix, (iii) unlike traditional LK the computational cost is invariant to the number of filters and as a result far more efficient, and (iv) this approach can be extended to the inverse compositional form of the LK algorithm where nearly all steps (including Fourier transform and filter bank pre-processing) can be pre-computed leading to an extremely efficient and robust approach to gradient descent image matching. Further, these computational savings translate to non-rigid object alignment tasks that are considered extensions of the LK algorithm such as those found in Active Appearance Models (AAMs).

**Index Terms**—Lucas & Kanade (LK), Fourier Domain, Illumination Invariance, Active Appearance Model (AAM).



## 1 INTRODUCTION

THE Lucas & Kanade (LK) algorithm [1] is the method of choice for direct image/object alignment<sup>1</sup>. As noted by Baker et al. [2] the modern LK algorithm is more than a single algorithm, but describes a family of algorithms suitable for numerous direct alignment tasks ranging from its canonical application of optical flow [1] to geometrically complex object alignment tasks such as those found in Active Appearance Models [3]. An extension to the LK algorithm that has gained considerable traction in the vision community over the last decade is the inverse compositional (IC) update [4]. LK-IC is of wide interest as it dramatically reduces the computational cost of the canonical LK algorithm with, in general, little to no loss in alignment performance. In fact it has been shown that in some circumstances the LK-IC algorithm outperforms the canonical LK algorithm [2]. The speedup realized from the IC update, however, is not equal across the LK family of algorithms. For the IC update to be of practical use it assumes that the Jacobian and Hessian matrices of the objective with respect to the warp parameters are static. A common objective employed to ensure this condition is the sum of squared distances (SSD).

A fundamental issue constantly faced by the SSD objective in nearly all forms of vision/learning is its poor performance in the presence of appearance variation. We take special care here

- S. Lucey is with the Commonwealth Scientific and Industrial Research Organisation (CSIRO), Brisbane, Australia. E-mail: [simon.lucey@csiro.au](mailto:simon.lucey@csiro.au)
- R. Navarathna and S.Sridharan are with the Speech, Audio, Image and Video Technology Lab, Queensland University of Technology, Australia. E-mail: [{r.navarathna, s.sridharan}@qut.edu.au](mailto:{r.navarathna, s.sridharan}@qut.edu.au)
- A. B. Ashraf is with the Computational Breast Imaging Group at the University of Pennsylvania School of Medicine, Philadelphia, USA. E-mail: [Ahmed.Ashraf@uphs.upenn.edu](mailto:Ahmed.Ashraf@uphs.upenn.edu)

<sup>1</sup>. It should be noted that there are, broadly speaking, two approaches to image/object alignment: (i) *feature based* approaches rely on abstracting the image/object by the geometric location of a set of chosen distinctive features, and (ii) *direct* approaches use all the pixels of interest. This paper is concerned with direct approaches to image/object matching.

to disambiguate “seen” versus “unseen” appearance variation. Numerous approaches have been proposed in LK literature for handling previously “seen” appearance variation. Typically these approaches attempt to learn an objective function from an ensemble of exemplars that model the space of appearance variation. Active Appearance Models are an example of this type of approach. However, a different approach needs to be entertained if one does not have examples of the space of appearance variation. One common approach for dealing with these unseen types of appearance changes is to apply a robust error function, instead of the canonical SSD objective. In the iterative setting of the LK algorithm, the employment of a robust error function results in solving a sequence of weighted linear SSD sub-problems. The weighting matrix adaptively changes each iteration as a function of the source image, model/template, and type of robust error function. Although being considerably less susceptible to outliers, than the canonical SSD objective, such functions are problematic when considering the IC update as they offer in practice no speedup due to the adaptively changing Jacobian and Hessian matrices.

Gaining invariance to previously unseen illumination variation while enjoying computationally efficient LK-style alignment performance through the IC update is the topic of central interest in this paper. We make a case for solving this problem by posing the LK family of algorithms in the 2D Fourier domain. The usefulness of such a framework becomes apparent if one explores the problem of applying the traditional LK algorithm simultaneously across a bank of filter image responses. Filter-banks have been shown to be useful in gaining invariance to spectral distortions such as those encountered in the presence of illumination variations on approximately Lambertian surfaces [5]. Unfortunately, in the spatial domain the complexity of the LK algorithm is a direct function of the number of filter banks being applied, greatly increasing the computational and memory requirements of image matching on raw pixels. Other areas of computer vision have dealt

with filter-bank computational and memory overhead through approximation methods [6], [7], [8], [9] (e.g., subsampling responses, feature selection, etc.).

In our work we instead argue to reformulate the problem by expressing the template/model and source image pixel intensities in the 2D Fourier domain. Through this reformulation one can instead solve the problem exactly, without any need for approximations, using an algorithm we refer to as Fourier LK (FLK). The FLK algorithm has a number of useful characteristics compared to the traditional LK algorithm namely: (i) computational complexity is *independent* of the number of filter banks employed, and (ii) the pre-processing step can be viewed within the canonical LK algorithm as solving a sequence of SSD sub-problems with a *fixed* weighting matrix. Our key contributions in this paper are as follows,

- Demonstrate that 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.
- Derive the inverse-compositional extension for the FLK algorithm and show that it significantly speeds up performance in comparison to the forwards-additive FLK formulation and has exactly the same computational load as that of the canonical inverse-compositional LK algorithm in the spatial domain (Section 6.2).
- Apply FLK to more complex object alignment tasks such as Active Appearance Models (AAMs) that can model object appearance variation. We demonstrate how the FLK philosophy can be incorporated into the estimation of the appearance basis, and how computationally efficient AAMs can be realized using the IC update and FLK framework within the project-out fitting strategy [3]. This result is one of the major contributions of this paper, as it allows one to obtain the favorable properties of multiple filter responses in the project-out algorithm without the need to explicitly compute the multiple responses.

Compared to our previous work [10], [11] this paper explores the application of the FLK framework across the entire family of LK algorithms ranging from canonical rigid image matching to more specialized non-rigid object alignment tasks such as those found in AAMs. Specifically, the Fourier “project-out” AAM fitting algorithm we present in this paper is substantially different to the Fourier AAM we first proposed in [11]. This new project-out extension of the FLK algorithm is extremely efficient as it can seamlessly incorporate an IC update with a static weighting matrix, unlike our previous Fourier AAM work which employed the computationally taxing “simultaneous” fitting strategy. Further, unlike our earlier work [10], [11], we give firm theoretical motivations for the choice/use of filter banks used within the FLK algorithm and describe qualitatively bounds on the types of filters that may be used within the framework.

### 1.1 General notation.

Vectors are always presented in lower-case bold (e.g.,  $\mathbf{a}$ ), Matrices are in upper-case bold (e.g.,  $\mathbf{A}$ ) and scalars in lower-case (e.g.  $a$ ). Images are expressed in capitalized form  $A$ .

Warp functions  $\mathcal{W}(\mathbf{x}_i; \mathbf{p}) = [\mathcal{W}_x(\mathbf{x}_i; \mathbf{p}), \mathcal{W}_y(\mathbf{x}_i; \mathbf{p})]^T$  will be used throughout this paper to denote a warping of the  $i$ th 2D coordinate vector  $\mathbf{x}_i = [x_i, y_i]^T$  by a warp parameter vector  $\mathbf{p} \in \mathcal{R}^P$ , where  $P$  is the number of warp parameters, back to the  $i$ th position in a fixed base coordinate system. The concatenated vector of all discrete positions in the base coordinate system shall be defined as  $\mathbf{x} = [x_1, \dots, x_D, y_1, \dots, y_D]^T$ , similarly the warp function across all the concatenated coordinates shall be described as  $\mathcal{W}(\mathbf{x}; \mathbf{p}) = [\mathcal{W}_x(\mathbf{x}_1; \mathbf{p}), \dots, \mathcal{W}_x(\mathbf{x}_D; \mathbf{p}), \mathcal{W}_y(\mathbf{x}_1; \mathbf{p}), \dots, \mathcal{W}_y(\mathbf{x}_D; \mathbf{p})]^T$ . 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  $A(\mathbf{p}) = [A(\mathcal{W}(\mathbf{x}_1; \mathbf{p})), \dots, A(\mathcal{W}(\mathbf{x}_D; \mathbf{p}))]^T$ . 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  $P \times D$  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 Section 5.4. Finally, we use the notation  $\|\mathbf{a}\|_{\mathbf{Q}}^2$  to represent the quadratic form  $\mathbf{a}^T \mathbf{Q} \mathbf{a}$ , and  $\mathbf{Q}$  is a symmetric, positive semi-definite weighting matrix.

### 1.2 Fourier notation.

The FLK framework uses key concepts from signal processing. A convolution operation is represented as the  $*$  operator. A  $\hat{\cdot}$  applied to any vector denotes the Discrete Fourier Transform (DFT) of an 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<sup>2</sup> 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 [12].

## 2 RELATED WORK

Of specific interest in this paper are computationally efficient approaches for image and object alignment within an LK inspired framework that can handle illumination variation. Earlier in this paper a distinction was made between previously

2. We should not that when dealing with a 2D real discrete signal  $\mathbf{X} \in \mathcal{R}^{N \times M}$ , where  $\mathbf{x} = \text{vec}(\mathbf{X})$ , the vectorized 2D Fourier transform is  $\hat{\mathbf{x}} = \text{vec}(\mathbf{F}_N \mathbf{X} \mathbf{F}_M^T) = (\mathbf{F}_M \otimes \mathbf{F}_N) \text{vec}(\mathbf{X}) = \mathbf{F} \mathbf{x}$ . Where  $\mathbf{F}_K$  is the 1-D Fourier transform matrix for a  $K$  sampled signal and  $\otimes$  is the Kronecker product operator.

“seen” and “unseen” appearance variation of an object. Given a number of examples (i.e. previously seen appearance) of an object under different illumination conditions, it is reasonably well understood now how to perform accurate LK inspired alignment. One of the earliest and notable efforts in this space is the work of Hager and Belhumeur [13], who demonstrated how to incorporate a linear appearance basis within the LK algorithm’s original SSD objective function. In their experiments, authors were able to perform alignment/tracking of an object with substantial illumination variation, if the offline examples from which the linear basis was formed cover that appearance basis space. Belhumeur and Kriegman [14] went on to show that for an object with arbitrary reflectance functions seen under arbitrary illumination conditions the image set of possible appearance is a convex cone. Further, for Lambertian objects with an arbitrary number of point light sources at infinity the image set is a convex polyhedral cone which can be determined from as few as three images. Matthews and Baker [3] demonstrated how the use of an arbitrary appearance basis can be incorporated into the LK-IC algorithm by proposing the “project-out” algorithm. Efficient object alignment results were demonstrated for non-rigid AAM face fitting tasks. Discriminative extensions to the LK algorithm have also been explored [15], [16], [17], [18] with limited success. A drawback to all these approaches, however, is the need to have “seen” offline examples of the object’s illumination variation.

For the case of previously “unseen” illumination variation, Black and Jepson [19] first proposed the employment of robust error functions within the LK framework and have shown them to be useful in the presence of illumination variation without the need for a priori knowledge. Theobald et al. [20] presented a study of the usefulness of robust-error functions for AAM fitting for dealing with previously unseen appearance variations. Although useful, this approach is problematic as it requires the need to reestimate the Hessian matrix at each iteration of fitting irrespective of the approach employed. This problem is particularly limiting for the inverse-compositional project-out algorithm as it dramatically slows down performance.

Robust error functions are not the only non-SSD objective function to be entertained in LK literature. Evangelidis and Psarakis [21] proposed the use of an enhanced correlation coefficient measure in a computationally efficient manner using an IC update. The approach, however, was limited to only giving invariance to photometric gain and bias distortions. Dowson and Bowden [22] have previously employed a mutual information (MI) measure within an LK framework. Their approach exhibited superior performance to a SSD objective under varying illumination, however, could not easily be extended to an IC form as the Jacobian and Hessian matrices change every iteration when employing an MI objective. An adhoc solution, however, was proposed by the authors by artificially fixing the Jacobian and Hessian matrices. Recently, Tzimiropoulos et al. [23] entertained the idea of using a correlation-based approach to handle non-uniform illumination variations. This recent work shares common themes with our proposed approach, as they propose to perform LK-style fitting

on filter responses rather than raw pixels. The approach is also computationally efficient as it is able to borrow upon LK inspired IC fitting. A drawback to both the MI [22] and correlation based [23] approaches, however, is that they are both concerned solely with image rather than object alignment. Extending the MI [22] and correlation [23] objectives to tasks where one also knows partially how an object’s appearance can vary (e.g. AAMs) remains at the moment non-trivial.

### 3 LK FITTING

The LK algorithm [1], [2] attempts to find the parametric warp  $\mathbf{p}$  that minimizes the SSD between a template image  $T$  and a warped source image  $I$ , given by the following error term,

$$\arg \min_{\mathbf{p}} \| I(\mathbf{p}) - T(\mathbf{0}) \|^2 \quad (1)$$

where  $I(\mathbf{p})$  represents the warped input image using the warp specified by the parameters  $\mathbf{p}$ , while  $T(\mathbf{0})$  represents the unwarped template image. In the original work of Lucas and Kanade [1] the parameter vector  $\mathbf{p}$  represented translation, but in principle one may represent any other type of parametric warp such as affine [19], [2] or piece-wise affine [24].

The weighted LK algorithm proposed by Baker and Matthews [25] attempts to find a warp that minimizes the following error,

$$\arg \min_{\mathbf{p}} \| I(\mathbf{p}) - T(\mathbf{0}) \|^2_{\mathbf{Q}} \quad (2)$$

where  $\mathbf{Q}$  is a symmetric, positive semi-definite weighting matrix. It is easy to see that when  $\mathbf{Q}$  is an identity matrix, the weighted and standard LK objective functions are equivalent. For this reason we shall refer to the standard LK algorithm herein as *Euclidean LK*.

Minimizing the error in Equation 2 is a non-linear optimization task. The weighted LK algorithm is able to find an effective solution to Equation 2, by iteratively linearizing  $I(\mathbf{p} + \Delta\mathbf{p})$  and refining the initial guess  $\mathbf{p} \leftarrow \mathbf{p} + \Delta\mathbf{p}$  at each iteration. The linearized objective function now takes the form,

$$\arg \min_{\Delta\mathbf{p}} \| I(\mathbf{p}) + \frac{\partial I(\mathbf{p})}{\partial \mathbf{p}} \Delta\mathbf{p} - T(\mathbf{0}) \|^2_{\mathbf{Q}} \quad (3)$$

which can be viewed as a quasi-Newton update. The explicit solution of  $\Delta\mathbf{p}$  that minimizes the linearized objective function is,

$$\Delta\mathbf{p} = \mathbf{H}^{-1} \frac{\partial I(\mathbf{p})}{\partial \mathbf{p}} \mathbf{Q} [T(\mathbf{0}) - I(\mathbf{p})] \quad (4)$$

where the pseudo Hessian matrix is defined by,

$$\mathbf{H} = \frac{\partial I(\mathbf{p})}{\partial \mathbf{p}} \mathbf{Q} \frac{\partial I(\mathbf{p})}{\partial \mathbf{p}}^T \quad (5)$$

The algorithm proceeds to update current warp parameters ( $\mathbf{p} \leftarrow \mathbf{p} + \Delta\mathbf{p}$ ) iteratively till convergence.

### 3.1 Inverse compositional fitting.

The canonical LK formulation presented in the previous section is sometimes referred to as the forwards additive (FA) algorithm [2]. A fundamental problem with the forwards additive approach is that it requires the re-estimation of the Hessian matrix (see Equation 5) since  $\frac{\partial I(\mathbf{p})}{\partial \mathbf{p}}$  must be re-computed at each iteration greatly impacting computational efficiency. Notably, Baker and Matthews [4] presented a computationally efficient extension to forwards additive LK which they refer to as the inverse compositional (IC) algorithm. Like the forwards additive algorithm the goal is to minimize the SSD objective function described in Equation 1. The approach differs, however, in that it linearizes  $T(\Delta \mathbf{p})$  rather than  $I(\mathbf{p} + \Delta \mathbf{p})$  resulting in the following linearized objective function,

$$\arg \min_{\Delta \mathbf{p}} \left\| I(\mathbf{p}) - T(\mathbf{0}) - \frac{\partial T(\mathbf{0})}{\partial \mathbf{p}} \Delta \mathbf{p} \right\|^2 \quad (6)$$

Since  $\frac{\partial T(\mathbf{0})}{\partial \mathbf{p}}$  needs only to be computed once, irrespective of the current value of  $\mathbf{p}$ , one can then solve this linearized objective function,

$$\Delta \mathbf{p} = \mathbf{B} [I(\mathbf{p}) - T(\mathbf{0})] \quad (7)$$

where  $\mathbf{B}$  can be completely pre-computed as,

$$\mathbf{B} = \mathbf{H}_{ic}^{-1} \frac{\partial T(\mathbf{0})}{\partial \mathbf{p}} \mathbf{Q} \quad (8)$$

and,

$$\mathbf{H}_{ic} = \frac{\partial T(\mathbf{0})}{\partial \mathbf{p}} \mathbf{Q} \frac{\partial T(\mathbf{0})}{\partial \mathbf{p}}^T \quad (9)$$

The current warp parameters are iteratively updated by the inverse (as we want to update the source image not the template) of the warp update  $\mathbf{p} \leftarrow \mathbf{p} \circ \Delta \mathbf{p}^{-1}$ . The operation  $\circ$  represents the composition of two warps which, in the case of an affine warp, can easily be represented as a matrix multiplication. For more complicated warps, such as the triangulated mesh commonly employed in AAMs, approximations can be employed [3] so as to ensure a compositional framework can be entertained. Due to its large computational advantage over the canonical FA algorithm, the IC form of the LK family of algorithms are of primary interest in this paper.

### 3.2 Robust error function.

It is well known the SSD objective is problematic in the presence of outliers. Robust error functions are designed to dampen the effect of these outliers. A common form used for LK style fitting to an image template  $T$  is,

$$\arg \min_{\mathbf{p}} \eta \{I(\mathbf{p}) - T(\mathbf{0})\} \quad (10)$$

where  $\eta\{\mathbf{t}\} = \sum_{d=1}^D \rho(t_d)$  and  $\mathbf{t}$  in this instance is an arbitrary vector with the same dimensionality  $D$  as the image template. It is the function  $\rho()$  that is commonly referred to as the robust error function.

#### 3.2.1 Robust IC fitting.

Applying an IC fitting strategy to the objective in Equation 10 results in an objective that is very similar in form to the traditional IC strategy using an SSD objective,

$$\arg \min_{\Delta \mathbf{p}} \eta \{I(\mathbf{p}) - T(\mathbf{0}) - \frac{\partial T(\mathbf{0})}{\partial \mathbf{p}} \Delta \mathbf{p}\} \quad (11)$$

where we iteratively solve for the warp update  $\Delta \mathbf{p}$  such that  $\mathbf{p} \leftarrow \mathbf{p} \circ \Delta \mathbf{p}^{-1}$ . The quasi-Newton update becomes,

$$\Delta \mathbf{p} = \mathbf{H}_{\eta}^{-1} \frac{\partial T(\mathbf{0})}{\partial \mathbf{p}} \nabla \eta \{I(\mathbf{p}) - T(\mathbf{0})\} \quad (12)$$

where  $\nabla \eta \{I(\mathbf{p}) - T(\mathbf{0})\}$  denotes the first derivative of the robust objective with respect to the error image. We define the pseudo-Hessian  $\mathbf{H}_{\eta}$  as,

$$\mathbf{H}_{\eta} = \frac{\partial T(\mathbf{0})}{\partial \mathbf{p}} \nabla^2 \eta \{I(\mathbf{p}) - T(\mathbf{0})\} \frac{\partial T(\mathbf{0})}{\partial \mathbf{p}}^T \quad (13)$$

where  $\nabla^2 \eta \{I(\mathbf{p}) - T(\mathbf{0})\}$  denotes the second derivative of the robust objective with respect to the error image. It is interesting to note that when we assume a canonical SSD objective  $\eta\{\mathbf{t}\} = \|\mathbf{t}\|^2$  then  $\nabla^2 \eta\{\mathbf{t}\} = \mathbf{I}$  irrespective of the value  $\mathbf{t}$  allowing  $\mathbf{H}_{\eta}$  to remain static. Unfortunately, for the general case  $\nabla^2 \eta\{\mathbf{t}\}$  is not static; thus dramatically affecting the efficiency of the IC algorithm when employing a robust error function since the Hessian must be re-computed at each iteration. It is this dilemma which is the central motivation for our paper.

#### 3.2.2 Choice of robust error function.

A number of functions for  $\rho()$  have been espoused in LK literature [25] such as Geman-McLure and Huber functions. Recently, Theobald et al. [20] demonstrated empirically that employing a decaying exponential,

$$\rho(t) = 1 - \exp(-\sigma \|t\|^2) \quad (14)$$

gave superior results in LK-style AAM fitting compared to these more traditional functions. This is the form of the robust error function that will be used throughout this paper. The decay factor  $\sigma$  is tuned through a cross-validation procedure.

## 4 AAM FITTING

Active appearance models (AAMs) [26], [3] are usually constructed from a set of training images with the AAM mesh vertices hand-labeled on them [26]. 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 [26]. The shape  $\mathbf{s}$  of an AAM is described by a 2D triangulated mesh. The 2D shape  $\mathbf{s} = (x_1, y_1, \dots, x_v, y_v)^T$  can be represented as a base shape  $\mathbf{s}_0$  plus a linear combination of  $P$  shape vectors  $\mathbf{s}_i$ :

$$\mathbf{s} = \mathbf{s}_0 + \sum_{i=1}^P p_i \mathbf{s}_i \quad (15)$$

where  $\mathbf{p} = [p_1, \dots, p_P]^T$  is the shape parameter vector. Camera parameters are also appended to the vector  $\mathbf{p}$  to model

the rigid variation of the object. 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  $A_\lambda(\mathbf{0})$  can be represented as a mean appearance  $A_0(\mathbf{0})$  plus a linear combination of  $K$  orthonormal appearance vectors  $A_j(\mathbf{0})$ :

$$\begin{aligned} A_\lambda(\mathbf{0}) &= A_0(\mathbf{0}) + \sum_{j=1}^K \lambda_j A_j(\mathbf{0}) \\ &= A_0(\mathbf{0}) + \mathbf{A}\boldsymbol{\lambda} \end{aligned} \quad (16)$$

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

#### 4.1 AAM fitting.

A number of approaches have been proposed in literature for fitting AAMs [26], [3]. The most notable and popular of these variants are approaches based on the LK algorithm [3]. In this approach one can pose AAM fitting as minimizing the following objective function:

$$\arg \min_{\mathbf{p}, \boldsymbol{\lambda}} \| I(\mathbf{p}) - A_0(\mathbf{0}) - \mathbf{A}\boldsymbol{\lambda} \|^2_{\mathbf{Q}} \quad (17)$$

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 17 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 17 is difficult to solve as there is a non-linear relationship between the shape  $\mathbf{p}$ , and appearance  $\boldsymbol{\lambda}$  parameters. A key insight, stemming from Lucas & Kanade [1], 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. For completeness the robust error function (Section 3.2) form can be equally applied to the problem of AAM fitting such that,

$$\arg \min_{\mathbf{p}, \boldsymbol{\lambda}} \eta \{ I(\mathbf{p}) - A_\lambda(\mathbf{0}) \} . \quad (18)$$

However, for the purposes of this section we will entertain the weighted Euclidean form as it is this form that is of immediate interest to our FLK formulation.

##### 4.1.1 Simultaneous algorithm.

The simultaneous algorithm [3] linearizes the objective function in Equation 17 such that:

$$\arg \min_{\Delta \mathbf{p}, \Delta \boldsymbol{\lambda}} \| I(\mathbf{p}) - A_\lambda(\mathbf{0}) - \frac{\partial A_\lambda(\mathbf{0})}{\partial \mathbf{p}} \Delta \mathbf{p} - \mathbf{A} \Delta \boldsymbol{\lambda} \|^2_{\mathbf{Q}} . \quad (19)$$

Instead of solving for the shape  $\mathbf{p}$  and appearance  $\boldsymbol{\lambda}$  parameters directly, through the linearization step in 19 we

iteratively solve for the updates  $\Delta \mathbf{p}$  and  $\Delta \boldsymbol{\lambda}$ . The objective function in Equation 19 takes advantage of a computationally efficient inverse compositional update for the warp parameters (described previously in Section 3.1). 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:

$$\begin{bmatrix} \Delta \mathbf{p} \\ \Delta \boldsymbol{\lambda} \end{bmatrix} = \mathbf{H}_{sim}^{-1} \mathbf{J}_{sim}^T \mathbf{Q} [I(\mathbf{p}) - A_\lambda(\mathbf{0})] \quad (20)$$

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

$$\mathbf{J}_{sim} = \begin{bmatrix} \frac{\partial A_\lambda(\mathbf{0})}{\partial \mathbf{p}} \\ \mathbf{A} \end{bmatrix} . \quad (21)$$

Empirically, the simultaneous algorithm has been noted to have excellent fitting performance compared to other LK inspired methods to AAM fitting. A major problem, however, with the simultaneous algorithm occurs with respect to computational efficiency. Specifically, even though an IC strategy is being employed (i.e. linearizing the model rather than the source image with respect to the warp update) a consequence of the appearance update step  $\boldsymbol{\lambda} \leftarrow \boldsymbol{\lambda} + \Delta \boldsymbol{\lambda}$  the appearance image  $A_\lambda(\mathbf{0})$ , Jacobian matrix  $\mathbf{J}_{sim}$ , and Hessian matrix  $\mathbf{H}_{sim}$  must be re-estimated at each iteration.

##### 4.1.2 Project-out algorithm.

The project-out algorithm [3] proposed by Matthews and Baker circumvents the computational limitations of the simultaneous algorithm by attempting to “project-out” appearance variation,

$$\arg \min_{\Delta \mathbf{p}} \| I(\mathbf{p}) - A_0(\mathbf{0}) - \frac{\partial A_0(\mathbf{0})}{\partial \mathbf{p}} \Delta \mathbf{p} \|^2_{\mathbf{Q}_\perp} \quad (22)$$

where  $\mathbf{Q}_\perp$  is the modified weight matrix where the appearance basis  $\mathbf{A}$  has been projected out,

$$\mathbf{Q}_\perp = \mathbf{Q} - (\mathbf{A}\mathbf{A}^T)\mathbf{Q}(\mathbf{A}\mathbf{A}^T) . \quad (23)$$

Equation 22 is an approximation to the simultaneous algorithm in Equation 19 as it is no longer updating the appearance template  $A_\lambda$  (note  $A_0$  is instead used, which is the mean appearance template).

The computational advantages of inverse compositional inspired fitting are readily apparent for the project-out algorithm. Specifically, one can solve the objective function in Equation 22 such that,

$$\Delta \mathbf{p} = \mathbf{B}_{po} [I(\mathbf{p}) - A_0(\mathbf{0})] \quad (24)$$

where  $\mathbf{B}_{po}$  can be completely pre-computed and the current warp parameters are iteratively updated by the inverse (as we want to update the source image not the template) of the warp update. The update matrix is defined as,

$$\mathbf{B}_{po} = \mathbf{H}_{po}^{-1} \mathbf{J}_{po}^T \mathbf{Q}_\perp \quad (25)$$

where the pseudo project-out Hessian matrix is  $\mathbf{H}_{po} = \mathbf{J}_{po}^T \mathbf{Q}_\perp \mathbf{J}_{po}$ . The project-out Jacobian matrix is defined as  $\mathbf{J}_{po} = \frac{\partial A_0(\mathbf{0})}{\partial \mathbf{p}}$ . For the project-out algorithm the Jacobian

and Hessian matrices remain static across all iterations, thus allowing  $\mathbf{B}_{po}$  to remain static. To gain an insight into the speedup that is afforded by pre-computing  $\mathbf{B}_{po}$ , implementations of project-out AAM face fitting have been reported in literature [3] running at a 200 fps on a modern PC (Intel core dual 3.0 GHz).

## 4.2 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:

$$\arg \max_{\mathbf{A}} \text{tr}(\mathbf{A}^T \mathbf{V} \mathbf{C} \mathbf{V}^T \mathbf{A}) \text{ subject to } \mathbf{A}^T \mathbf{A} = \mathbf{I} \quad (26)$$

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

## 5 FITTING WITH FILTER RESPONSES

The employment of filter banks as a pre-processing step in many tasks in vision involving illumination variations is motivated by two widely accepted assumptions about human vision: (i) human vision is mostly sensitive to scene reflectance and mostly insensitive to the illumination conditions, and (ii) human vision responds to local changes in contrast rather than to global brightness levels [27]. These two assumptions are closely related since local contrast is a function of reflectance. A natural way to encode local contrast is through the employment of a bank of filters that encode local intensity differences at different orientations and scales. Many types of filters are possible for this task, common candidates used in vision applications are 2D Gabor functions; but any type of filter that encodes relative intensity differences across many different orientations and scales is suitable.

### 5.1 Spatial domain.

We can reformulate the LK algorithm to entertain fitting across multiple linear filter responses. This error functions can be written as,

$$\arg \min_{\mathbf{p}} \left\| \{\mathbf{g}_i * I(\mathbf{p})\}_{i=1}^M - \{\mathbf{g}_i * T(\mathbf{0})\}_{i=1}^M \right\|^2 \quad (27)$$

Where  $\mathbf{g}_i$  is  $i$ -th filter with  $M$  filters in total, while  $\{\cdot\}_{i=1}^M$  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.

#### 5.1.1 Computational cost.

As pointed out by [28], [8], [7] a particular problem with Equation 27 is the inherently large memory and computational overheads required for representing images in this over-complete filter response domain. The main fundamental problems when applying to the LK framework are:

- 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.
- The number of columns in the Jacobian  $\mathbf{J}$  matrix increases from  $PD$  to  $PMD$ , where  $P$  is the number of warp parameters. For the special case of the simultaneous algorithm  $P$  refers to the number of warp & appearance parameters.
- The computational cost for building Hessian  $\mathbf{H}$  matrix increases from  $P^2D$  to  $P^2MD$ .

As a result of these computational overheads, the idea of doing object alignment with even a modest number of Gabor filter banks (e.g., 9 scales times 8 orientations, i.e.  $M = 72$ , as employed in [7]) becomes prohibitively expensive and impractical when employing the forwards additive algorithm. Even for the inverse compositional algorithm, where the Jacobian and Hessian matrices can be pre-computed to form  $\mathbf{B}$ , the additional cost of estimating the overcomplete image representation  $\{\mathbf{g}_i * I(\mathbf{p})\}_{i=1}^M$  and  $M$ -fold increase in the size of the pre-computed matrix  $\mathbf{B}$  remains. 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 [7], (ii) employing filter responses at certain fiducial positions within the image [6], (iii) the employment of feature selection methods to select the most discriminative filter responses [8], 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 [9].

### 5.2 Fourier LK.

It is elementary to show that the error in Equation 27 can equivalently be written as,

$$\arg \min_{\mathbf{p}} \sum_{i=1}^M \left\| \mathbf{g}_i * [I(\mathbf{p}) - T(\mathbf{0})] \right\|^2 \quad (28)$$

Exploiting the fact that convolution becomes a Hadamard (i.e., element-by-element) product in the Fourier domain, and employing Parseval's relation [12] (energy content is preserved as we move from the spatial to the Fourier domain), we may write the error in Equation 28 as follows,

$$\arg \min_{\mathbf{p}} \left\| \hat{I}(\mathbf{p}) - \hat{T}(\mathbf{0}) \right\|_{\mathbf{S}}^2 \quad (29)$$

where,

$$\mathbf{S} = \sum_{i=1}^M \text{diag}(\hat{\mathbf{g}}_i)^T \text{diag}(\hat{\mathbf{g}}_i) \quad (30)$$

and  $\hat{I}, \hat{T}, \hat{\mathbf{g}}_i$  are the 2D Fourier transforms of vectorized  $I, T, \mathbf{g}_i$  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 by subsuming Equation 29 into the weighted LK objective function,

$$\arg \min_{\mathbf{p}} \left\| I(\mathbf{p}) - T(\mathbf{0}) \right\|_{\mathbf{F}^T \mathbf{S} \mathbf{F}}^2 \quad (31)$$

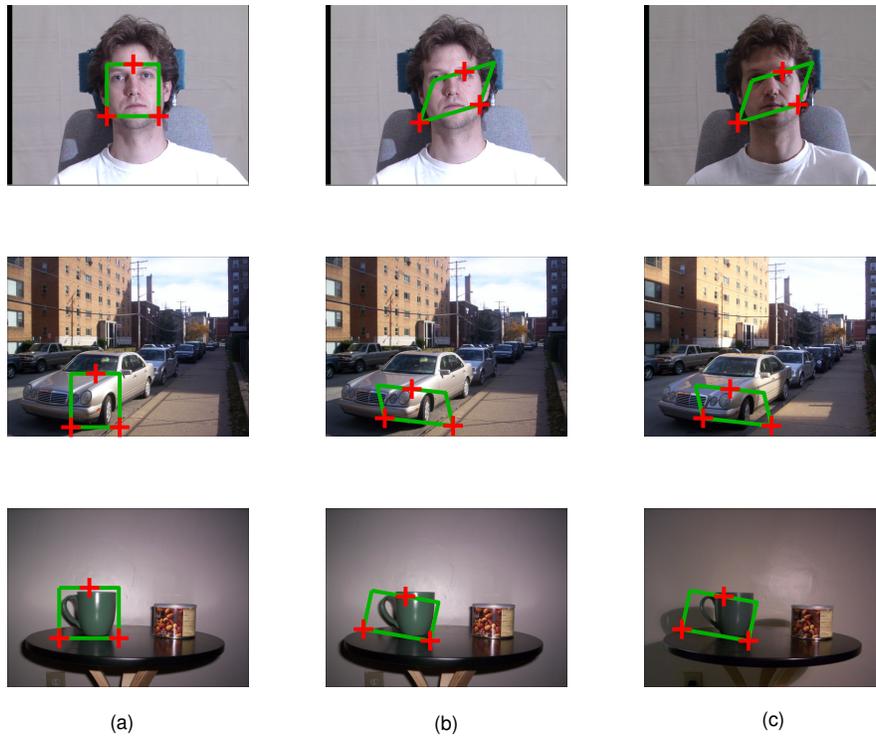


Fig. 1. (a) A template centered around the object of interest with three canonical points shown as red pluses. (b) Sample warps defined by the random perturbation of the three canonical points. (c) Sample warps shown on an image that has different illumination conditions as compared to that of the template for various objects.

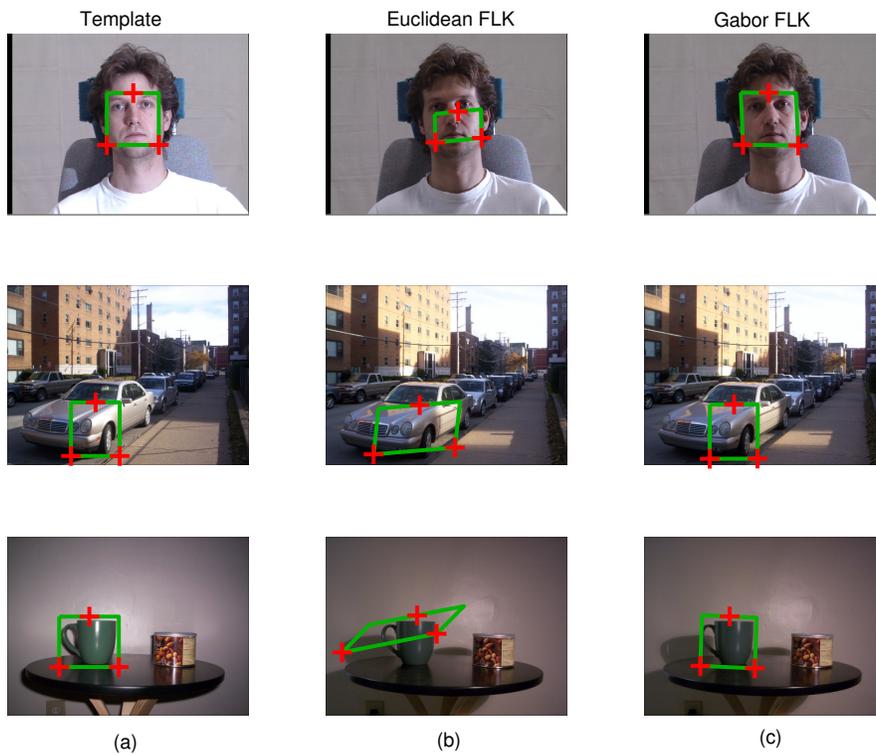


Fig. 2. Example fittings for the warps shown in Figure 1 in the presence of illumination. (a) Shows the template image. (b) Euclidean FLK is off-target in the presence of illumination variability in the source image. (c) Gabor FLK is robust in the presence of illumination variability in the source image.

### 5.2.1 A link to weighted LK.

By casting the LK algorithm in the Fourier domain, we have shown that doing alignment on linear filter preprocessed images is equivalent to the weighted LK algorithm with a weighting matrix  $\mathbf{Q} = \mathbf{F}^T \mathbf{S} \mathbf{F}$  where  $\mathbf{S}$  (Equation 30) is determined by the choice of filters being used. Deriving the weighting matrix  $\mathbf{Q}$  directly from a bank of filters is advantageous as they naturally encode redundancies present in the human visual system. Moreover we have described a practical method for applying a richer set of weighting matrices ( $\mathbf{Q} = \mathbf{F}^T \mathbf{S} \mathbf{F}$  is full rank and non-diagonal) in the spatial domain by just applying a diagonal weighting matrix  $\mathbf{S}$  in the Fourier domain. We can also see that FLK and LK inspired fitting strategies become equivalent when  $\mathbf{S} = \mathbf{I}$  since  $\mathbf{F}^T \mathbf{F} = \mathbf{I}^3$  which is a direct result of Parseval’s relation [12].

A result of particular importance is with regard to inverse compositional fitting, as it was demonstrated in Equation 8 that the role of  $\mathbf{Q}$  can be completely subsumed into the pre-computed update matrix  $\mathbf{B}$ . Of particular note is that since  $\mathbf{Q}$  contains the Fourier transform operation on both the template and image pre-computed, there is no online filtering step per iteration of the Fourier LK algorithm when using an inverse compositional update. As a result the online computational cost of the inverse compositional Fourier LK algorithm is identical to the canonical inverse compositional LK algorithm. The computational dividends accrued by the inverse compositional formulation of the FLK algorithm are evident in Table 1. If  $P$  and  $D$  represent the number of warp parameters and the number of pixels in the template image respectively the per-iteration computational complexity of forwards additive FLK is  $O(P^3 + P^2D + PD \log D)$ , while that of inverse compositional FLK is  $O(P^2 + PD)$ . We would like to emphasize that in both the cases, the complexity is independent of the number of filters,  $M$ .

### 5.3 Fourier simultaneous and project-out.

By establishing a link to weighted LK through the weighting matrix  $\mathbf{Q} = \mathbf{F}^T \mathbf{S} \mathbf{F}$  it becomes obvious that other extensions of the LK algorithmic family that employ a weighted Euclidean distance become applicable. Of particular note are the “simultaneous” and “project-out” fitting algorithms reviewed in Section 4 due to their applicability to AAMs.

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. For the project-out algorithm, however, the entire role of  $\mathbf{Q}$  and its Fourier transform  $\mathbf{F}$  can be completely pre-computed incurring *no* additional computational cost. In fact, the computational cost of the FLK project algorithm is identical to the online FLK inverse compositional algorithm cost outlined in Table 1(c). An additional cost is incurred for the project

algorithm in the pre-computation step as the appearance basis must be “projected out” to form the weight matrix  $\mathbf{Q}_\perp$ . This result is one of the major contributions of this paper, as it allows one to obtain the favorable properties of multiple filter responses in the project-out algorithm without the need to explicitly compute those multiple responses.

### 5.4 Image gradient attenuation.

A key element of our proposed objective in Equation 27 is that to solve the objective using an LK strategy requires the linearization of,

$$\mathbf{g}_i * T(\Delta \mathbf{p}) \approx \mathbf{g}_i * T(\mathbf{0}) + \frac{\partial \mathbf{g}_i * T(\mathbf{0})}{\partial \mathbf{p}} \Delta \mathbf{p} \quad (32)$$

where,

$$\mathbf{g}_i * \frac{\partial T(\mathbf{0})}{\partial \mathbf{p}} = \frac{\partial \mathcal{W}(\mathbf{x}; \mathbf{0})}{\partial \mathbf{p}} \frac{\partial [\mathbf{g}_i * T(\mathbf{0})]}{\partial \mathcal{W}(\mathbf{x}; \mathbf{0})} \quad (33)$$

such that  $\frac{\partial \mathcal{W}(\mathbf{x}; \mathbf{0})}{\partial \mathbf{p}}$  is the Jacobian of the warp function. In practice we evaluate the Jacobian of the template image with respect to the  $x$ - and  $y$ - coordinates using image gradients,

$$\frac{\partial [\mathbf{g}_i * T(\mathbf{0})]}{\partial \mathcal{W}(\mathbf{x}; \mathbf{0})} = \begin{bmatrix} \text{diag}\{\mathbf{g}_x * \mathbf{g}_i * T(\mathbf{0})\} \\ \text{diag}\{\mathbf{g}_y * \mathbf{g}_i * T(\mathbf{0})\} \end{bmatrix} \quad (34)$$

where  $\mathbf{g}_x$  and  $\mathbf{g}_y$  are the 2D gradient filters in the  $x$ - and  $y$ - directions respectively. An obvious limitation of our proposed formulation is that estimation of image gradients is itself a filter response. Typically the  $\mathbf{g}_x$  and  $\mathbf{g}_y$  gradient filters are horizontally and vertically oriented edge filters of a pre-defined scale. In nearly all cases we can, by definition, assume that  $\mathbf{g}_x$ ,  $\mathbf{g}_y$  and  $\mathbf{g}_i$  are bandpass filters (e.g. Gabor filters). An immediate issue when attempting to convolve two bandpass filters with one another (e.g.  $\mathbf{g}_x * \mathbf{g}_i$  or  $\mathbf{g}_y * \mathbf{g}_i$ ) is passband mismatch. Passband is the portion of the frequency spectrum that is transmitted with minimum relative loss. When the passband of two bandpass filters are mismatched then substantial attenuation can occur across the passbands of both filters when they are combined. In vision terms this results in “image gradient attenuation” as we obtain a largely useless linearization of  $\mathbf{g}_i * T(\Delta \mathbf{p})$ . Fortunately, however, in our proposed approach we never use a single pre-processing filter instead choosing to use a bank of  $M$  filters across multiple scales and orientations, or in signal processing terms multiple passband sizes and locations. We have found empirically using this strategy obtains a good linearization of  $\{\mathbf{g}_i * T(\Delta \mathbf{p})\}_{i=1}^M$  in nearly all circumstances with very little attenuation, regardless of the choice of  $\mathbf{g}_x$  and  $\mathbf{g}_y$ .

## 6 MULTIPLE EXPERIMENTS

The experiments were conducted in two phases, (a) image alignment in LK framework (b) AAM fitting. To compare the algorithms in terms of their robustness to illumination variation, the quantitative experiments were conducted with the MultiPIE face image dataset [29]. The MultiPIE database consisted of 19 illumination conditions (i.e., 18 variations of flash firing and without flash) with a range of facial expression including *neutral, smiles, surprise, squints, disgust*

3. 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.

No.	Step	Complexity
1	Warp $I$ with $\mathbf{p}$ to compute $I(\mathbf{p})$	$O(PD)$
2	Compute the error image: $I(\mathbf{p}) - T(\mathbf{0})$	$O(D)$
3	Evaluate the Jacobian $\mathbf{J} = \left(\frac{\partial I(\mathbf{p})}{\partial \mathbf{p}}\right)^T$	$O(PD)$
4	Compute FFT of the Jacobian	$O(PD \log D)$
5	Compute the Hessian	$O(P^2 D + PD)$
6	Invert the Hessian	$O(P^3)$
7	Compute FFT of the error image	$O(D \log D)$
8	Compute $(\mathbf{F}\mathbf{J})^T \mathbf{S}\mathbf{F}[I(\mathbf{p}) - T(\mathbf{0})]$	$O(PD + D)$
9	Compute $\Delta \mathbf{p}$	$O(P^2)$
10	Update $\mathbf{p}$ : $\mathbf{p} \leftarrow \mathbf{p} + \Delta \mathbf{p}$	$O(P)$
	<b>Total</b>	$O(P^3 + P^2 D + PD \log D)$

(a)

No.	Step	Complexity
1	Evaluate the Jacobian $\mathbf{J} = \left(\frac{\partial T(\mathbf{0})}{\partial \mathbf{p}}\right)^T$	$O(PD)$
2	Compute FFT for the Jacobian	$O(PD \log D)$
3	Compute the Hessian	$O(P^2 D + PD)$
4	Invert the Hessian	$O(P^3)$
5	Compute the matrix $\mathbf{B}$ (Eq. 8)	$O(P^2 D + PD + PD^2)$
	<b>Total</b>	$O(P^3 + P^2 D + PD^2 + PD \log D)$

(b)

No.	Step	Complexity
1	Warp $I$ with $\mathbf{p}$ to compute $I(\mathbf{p})$	$O(PD)$
2	Compute the error image: $I(\mathbf{p}) - T(\mathbf{0})$	$O(D)$
3	Compute $\Delta \mathbf{p} = \mathbf{B}[I(\mathbf{p}) - T(\mathbf{0})]$	$O(PD)$
4	Update $\mathbf{p}$ : $\mathbf{p} \leftarrow \mathbf{p} \circ \Delta \mathbf{p}^{-1}$	$O(P^2)$
	<b>Total</b>	$O(P^2 + PD)$

(c)

TABLE 1

Comparison of per-iteration complexity between the forwards additive and the inverse compositional formulations of the FLK algorithm.  $P$  and  $D$  represent the number of warp parameters and the number of pixels in the template image respectively, (a) per-iteration complexity for the FA algorithm, (b) complexity of pre-computations for the IC algorithm, and (c) per-iteration complexity for the IC algorithm.

and screams. Examples of this variation can be seen in Figure 3. All images were hand annotated with 68 points for AAM fitting and three points for LK image alignment. More details about the MultiPIE database can be found in [29]. An additional non-face image set (e.g. cars, cups, etc.) was also used for the LK image alignment task.

Throughout this section we will be comparing LK & 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. We shall refer to all variants of (i) and (ii) as Euclidean and Gabor Fourier LK (FLK). For all our experiments, 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.

### 6.1 Fitting in LK framework.

The experiments were similar to the methodology used in [2]. Given an image of an object (face, cup, car, etc.) we selected three canonical points indicated by red pluses in Figure 1(a). We then randomly perturbed these three points with additive white Gaussian noise of certain variance. The new location of these points describe an affine warp. Sample affine warps generated using this method are shown in Figure 1(b).

Figure 2 presents some visual examples of fitting performed by the two algorithms in the presence of illumination variability. In Figure 2(b) the Euclidean FLK is off target, while in Figure 2(c) Gabor FLK shows its robustness to illumination changes. For a quantitative assessment we compared the algorithms in terms of average convergence rates and average convergence frequency as described in the following subsection.

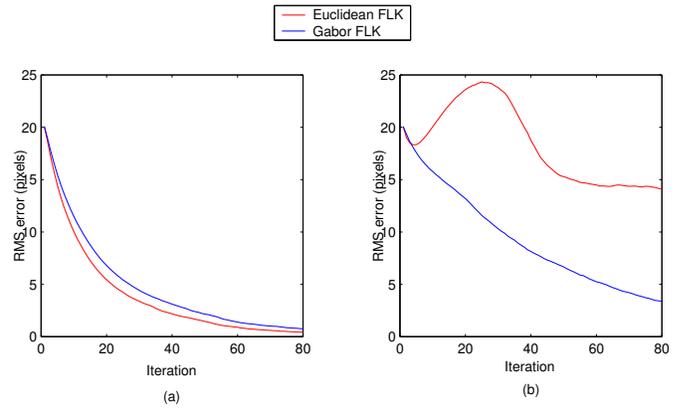


Fig. 4. Average convergence rates over 3000 randomly generated warps ranging from an initial RMS error of 10-35 pixels. (a) When the input and template images have the same illumination conditions, both algorithms perform equally well. (b) When the illumination of the input image changes, the Gabor FLK algorithm is still able to do the fitting, while the Euclidean LK fails.

#### 6.1.1 Average convergence rates.

We conducted fitting experiments on a large number (3000) of randomly generated warps. The initial RMS error for these warps ranged from 10-35 pixels. The RMS error plotted as a function of the iteration number is shown in Figure 4. When the lighting conditions for the template and the input image are the same, both algorithms have similar performance (Figure 4(a)). When the lighting conditions for the input image are changed, the Gabor FLK method is robust to the variability, while the Euclidean FLK algorithm diverges (Figure 4(b)).

#### 6.1.2 Average frequency of convergence.

To check the average frequency of convergence given an initial RMS error, we generated 3000 warps ranging from

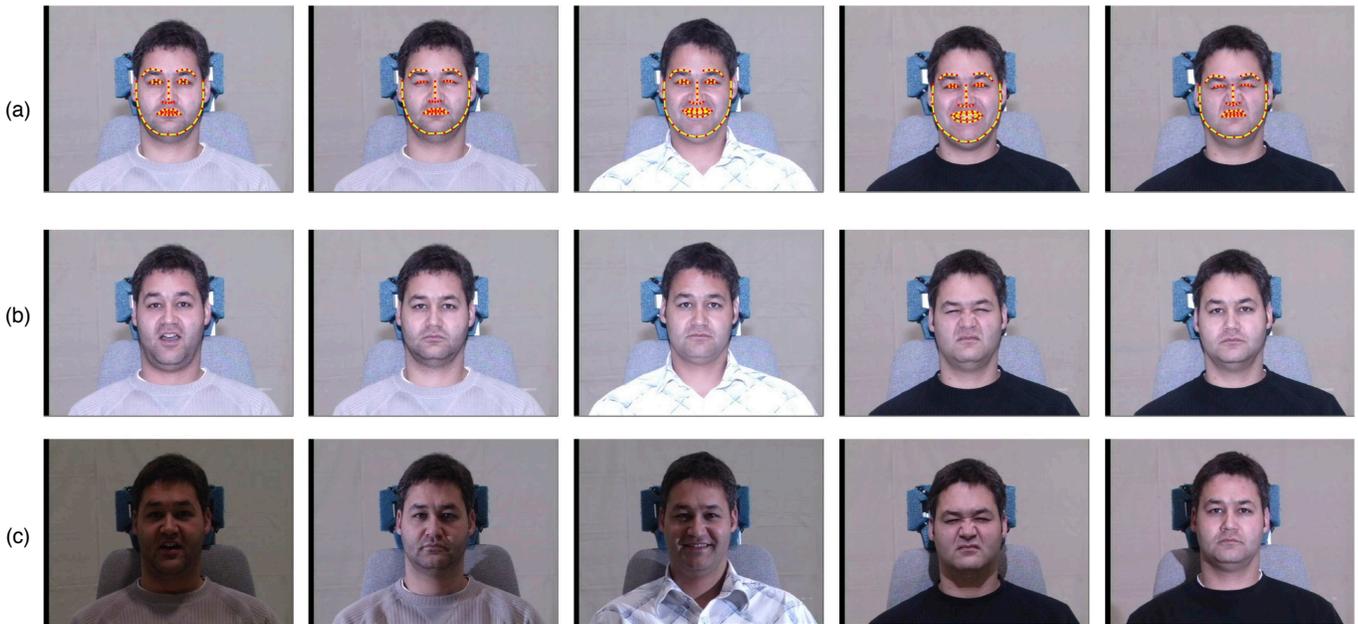


Fig. 3. (a) Five of the hand annotated (68 points) ground-truth images for a specific subject used to construct the AAM model. (b) Sample of testing images with the matched illumination condition. (c) Sample of testing images with the mismatched illumination condition.

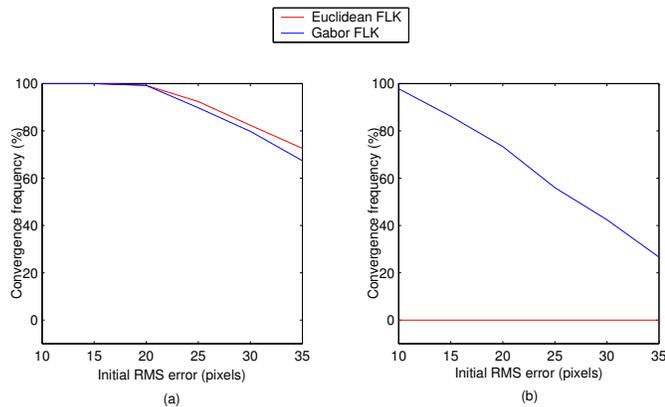


Fig. 5. Average frequency of convergence as a function of initial RMS. (a) When the input and template images have the same illumination conditions, both algorithms perform equally well. (b) When the illumination of the input image changes, the Gabor FLK algorithm clearly outperforms the Euclidean FLK for which the convergence frequency remains zero for all values of initial RMS error.

10-35 pixels. The algorithm was deemed to be converged if the final error was less than 5 pixels. We present the curves for convergence frequency, as a percentage, in Figure 5. Again the two algorithms have similar performance when the lighting conditions are the same for the input and template image (Figure 5(a)). When the illumination is mismatched, the Gabor FLK clearly outperforms the Euclidean FLK. This demonstrates the hypersensitivity of the conventional LK to illumination changes, while underscoring the robustness of Gabor FLK algorithm.

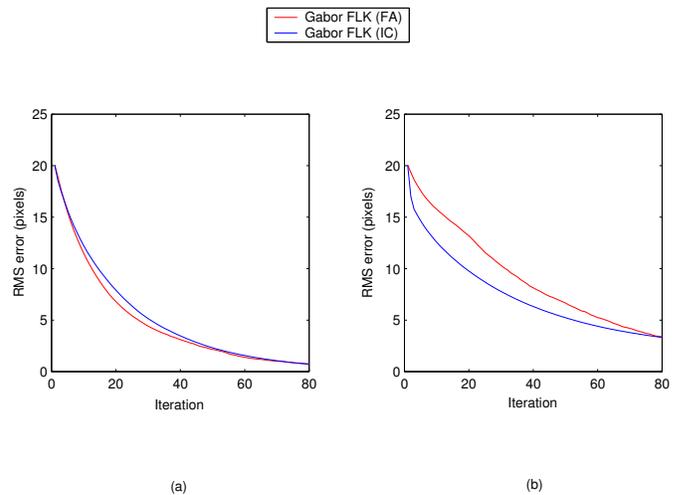


Fig. 6. Equivalence between of Gabor FLK forwards additive and inverse compositional algorithms for average convergence rates over 3000 randomly generated warps ranging from an initial RMS error of 10-35 pixels when: (a) the input and template images have the same illumination conditions, and (b) the input and template images have different illumination conditions.

### 6.2 Inverse compositional fitting.

The FLK results presented in the previous section adheres to the canonical form of the LK algorithm known as the forwards additive (FA) algorithm [2]. As discussed in Section 3.1 a fundamental problem with the forwards additive approach is that it requires the re-estimation of the Hessian matrix at each iteration greatly impacting computational efficiency. Notably, Baker and Matthews [4] proposed the inverse compositional

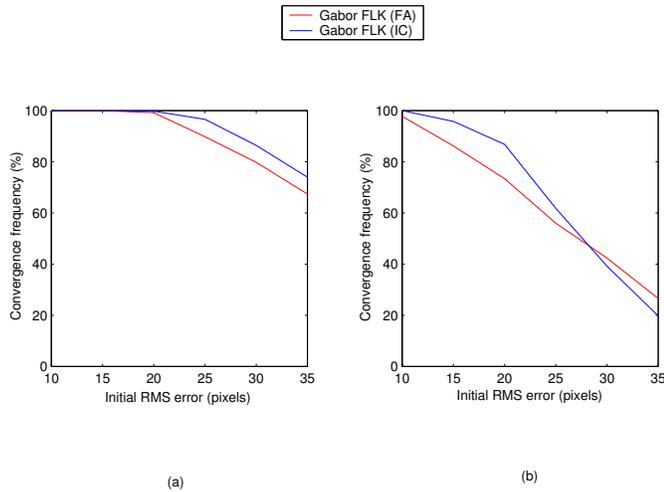


Fig. 7. Equivalence of Gabor FLK forwards additive (FA) and inverse compositional (IC) algorithms for average frequency of convergence as a function of initial RMS error when: (a) the input and template images have the same illumination conditions, and (b) the input and template images have different illumination conditions.

algorithm to circumvent this computational roadblock. The average convergence rates between the two formulations is demonstrated in Figure 6 and the frequency of convergence is presented in Figure 7 using a  $S$  matrix formed from Gabor filters.

### 6.3 AAM fitting.

Person specific AAM fitting algorithms were compared with the weighting matrices: (i)  $Q = I$ , and (ii)  $Q = F^T S F$  as in Section 6.1. We shall refer to all variants of (i) and (ii) as Euclidean Active Appearance Models (AAM) and Gabor Fourier Active Appearance Models (FAAM). The experiments were conducted using the simultaneous and project-out algorithms. In all our experiments the geometrically normalized AAM base template had an inter-ocular distance of 50 pixels. Person specific AAMs were chosen for our experiments over generic AAMs due to their superior registration accuracy and computational efficiency. 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. 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 [30].

#### 6.3.1 Simultaneous results.

Visual examples of fitting performance can be seen in Figures 8 and 9 for Euclidean AAM and Gabor FAAM respectively. Figure 10 depicts the average RMS mesh location

error against iterations for simultaneous variants of Euclidean AAMs and Gabor FAAMs for (a) matched and (b) mismatched illumination. Similarly, Figure 11 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 the Gabor FAAM.

#### 6.3.2 Project-out results.

Figure 12 depicts the average RMS mesh location error against iterations for project-out variants of Euclidean AAM and Gabor FAAM for (a) matched and (b) mismatched illumination. Figure 13 depicts the number of converged trials as a function of the RMS error threshold for (a) and (b). As expected, due to the approximation made by the project-out algorithm, results in Figures 12 and 13 are poorer than the simultaneous results depicted in Figures 10 and 11. In a similar fashion to the simultaneous results, however, (a) obtains almost identical performance to Euclidean AAM and Gabor FAAM. In the presence of substantial illumination mismatch (b) the Gabor FAAM outperforms Euclidean AAM by a substantial margin with no additional computational burden during online fitting.

#### 6.3.3 FAAM with a robust error function.

From the simultaneous and project-out experiments, we observed that Gabor FAAM obtained better alignment compared with the Euclidean AAM in the presence of illumination. We also evaluated the application of a robust error function, as described in Section 3.2. Figure 14 depicts the average RMS mesh location error against iterations for Gabor FAAM with and without a robust error function for: (a) matched and (b) mismatch illumination. Figure 15 depicts the number of converged trials as a function of the RMS error threshold for (a) and (b). For the case of matched illumination, (a) obtains almost identical performance with and without the robust error function. In the presence of substantial illumination mismatch, (b) both algorithms still perform almost identically as depicted in Figures 14 and 15. This result is of particular importance with respect to the project-out algorithm as the inclusion of the robust error function dramatically reduces computational performance.

## 7 TRACKING EXPERIMENTS

We conducted various tracking experiments on video sequences containing substantial variations in illumination over time using the Euclidean AAM and Gabor FAAM. An example of tracking sequence can be seen in Figure 16. 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 error from ground-truth can be seen in Figure 16 showing a substantial benefit to Gabor FAAM in person specific face tracking tasks. Visual examples of tracking performance in challenging environments are shown in Figure 17. All tracking results were obtained using a simultaneous fitting strategy.

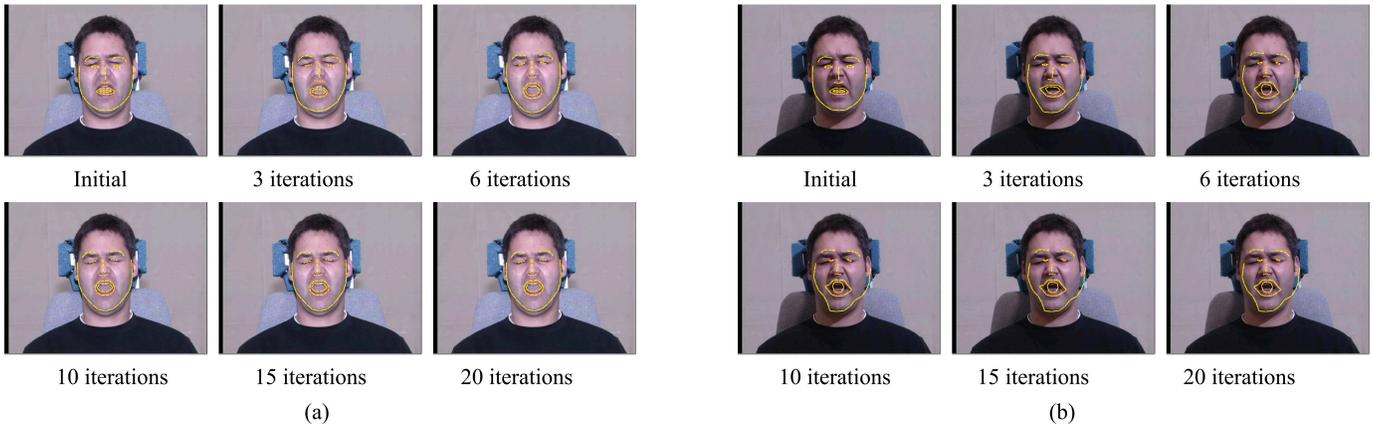


Fig. 8. Examples of tracking with the Euclidean AAM: (a) iteration frames with the matched illumination condition, and (b) iteration frames with the mismatched illumination condition.



Fig. 9. Examples of tracking with the Gabor FAAM: (a) iteration frames with the matched illumination condition, and (b) iteration frames with the mismatched illumination condition.

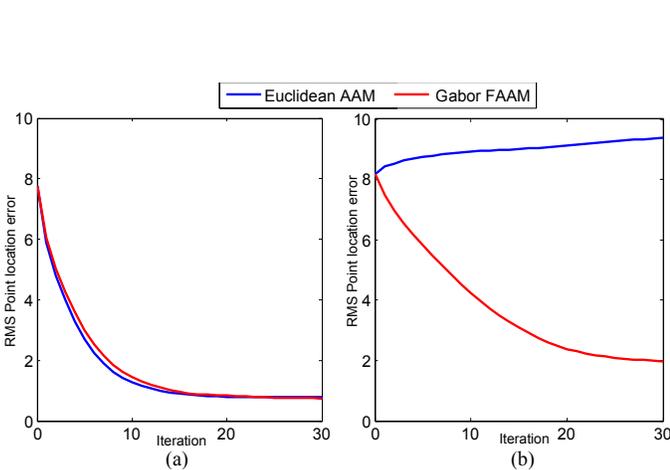


Fig. 10. Average convergence rates for the simultaneous algorithm. (a) When the input and training images have matched illumination conditions, both algorithms perform equally well. (b) When illumination is mismatched the Gabor FAAM algorithm is still able to perform well, while the Euclidean AAM algorithm diverges.

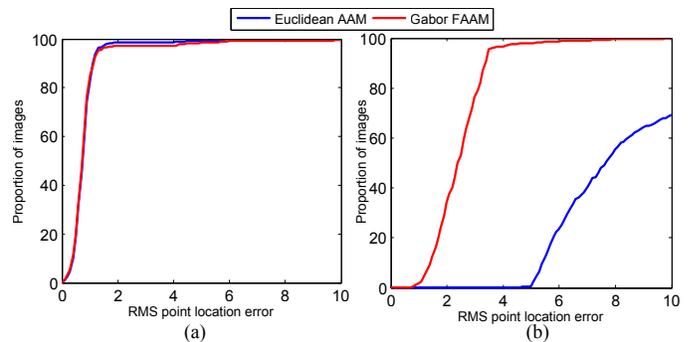


Fig. 11. Fitting performance curves for the simultaneous algorithm using Euclidean AAM and Gabor FAAM when: (a) the input and training images have matched illumination, (b) the input and training images have mismatched illumination.

## 8 DISCUSSION

Experimental results show that image alignment in Fourier domain using filter banks that encode relative local intensity (e.g. Gabor filters) clearly outperforms conventional unweighted SSD objectives. An interesting question arising from this result

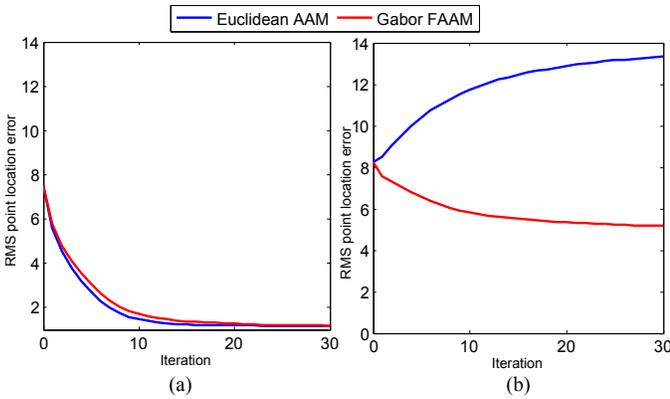


Fig. 12. Average convergence rates for the project-out algorithm. (a) When the input and training images have matched illumination conditions, both algorithms perform equally well. (b) When illumination is mismatched the Gabor FAAM algorithm is still able to do converge, while the Euclidean AAM algorithm diverges.

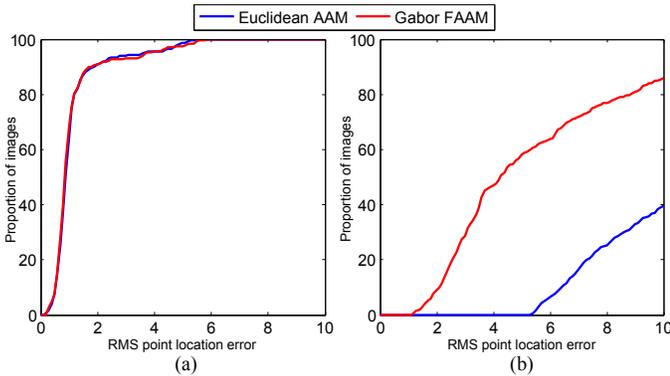


Fig. 13. Fitting performance curves for the project-out algorithm using Euclidean AAM and Gabor FAAM when: (a) the input and training images have matched illumination, and (b) the input and training images have mismatched illumination.

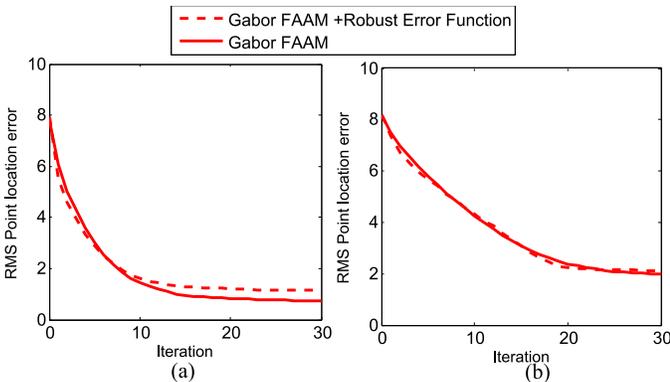


Fig. 14. Average convergence rates for Gabor FAAM and Gabor FAAM with a robust error function: Both algorithms perform almost identical for (a) match illumination conditions (b) mis-match illumination condition.

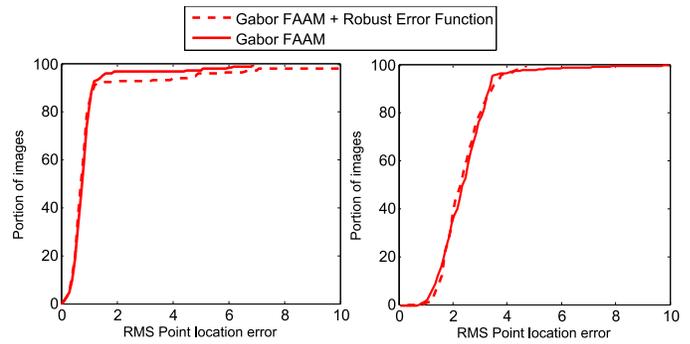


Fig. 15. Fitting performance curves for Gabor FAAM and Gabor FAAM with a robust error function : (a) when the input and training images have the same illumination, (b) when the input and training images have the mismatched illumination.

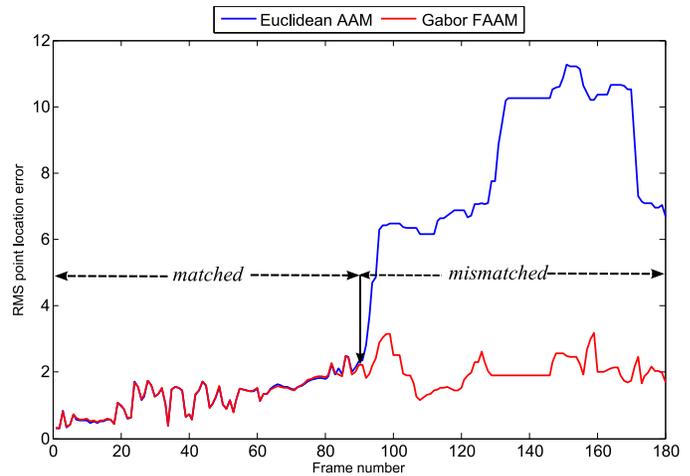
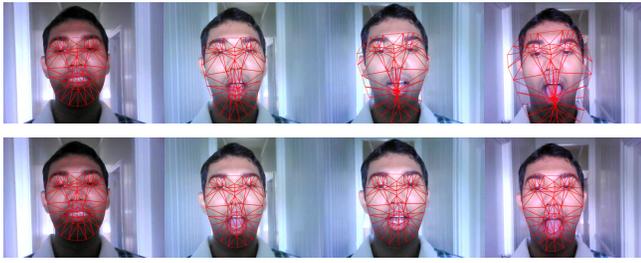


Fig. 16. Example of a tracking with Euclidean AAM and Gabor FAAM in a video sequence. Illumination is changing over time using the 3 different flashes. Euclidean AAM and the Gabor FAAM show similar results in the initial frames, but only Gabor FAAM showed good tracking results when the illumination changing over the time.

is: “Why does fitting across multiple filtered image responses improve alignment performance?”. Our work shows that the choice of the distance metric is crucial to the performance of an alignment algorithm, and linear filters provide a principled method to manipulate the distance metric through the weighting matrix  $S$  in the Fourier domain. A question for future work should perhaps be now, “What is the best weighting matrix  $S$  for my alignment problem?” rather than “What filter banks give best performance?”

## 9 CONCLUSION

In this paper we have demonstrated a novel extension to LK algorithm & AAM fitting algorithm which allows for them to be equivalently cast in the Fourier domain. This formulation allows us to interpret the joint alignment across multiple filter responses as a form of the weighted LK algorithm. Here, we have presented a method to do image & object alignment using



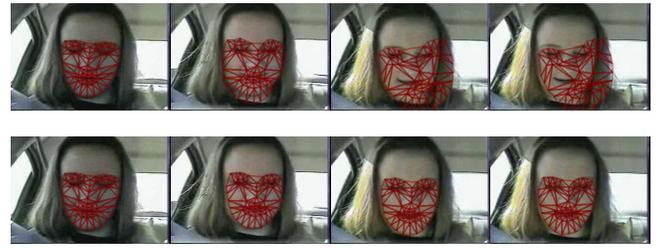
(a) Key frames taken from a video sequence of a person walking along a passage in a house.



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



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



(d) Tracking with sequence of image frames in a real-world automobile environment. Note: The video sequence was obtained from the AVICAR [31] database.

Fig. 17. Challenging examples of tracking in “real-world” scenarios where illumination conditions change dynamically in the sequence. Gabor FAAM in nearly all cases obtains better tracking results in comparison to Euclidean AAM. Top Row : tracking sequence with the Euclidean AAM. Bottom row: tracking sequence with the Gabor FAAM

a bank of filters that encode local relative intensity (e.g. Gabor filters). We have shown that doing image & object alignment in the high dimensional multiple filter response space is mathematically equivalent to doing alignment in a lower dimensional image intensity space, if appropriate weightings are applied in the Fourier domain. We have compared our Gabor FLK algorithm with the conventional Euclidean FLK and have shown that Gabor FLK outperforms Euclidean FLK in the presence of variabilities like illumination changes. We have demonstrated similar improvements with respect to more exotic LK style object alignment algorithms, specifically AAM fitting. We have extended our work in [10], [11] to person specific AAM fitting performance using project-out algorithm. For the project-out algorithm, the weighting matrix and its Fourier transform can be completely precomputed incurring no additional computational cost. This result is one of the major contributions of this paper, as it allows one to obtain the favorable properties of multiple filter responses in the project-out algorithm without the need to explicitly compute multiple responses.

## ACKNOWLEDGMENTS

Dr Simon Lucey is the recipient of an Australian Research Council Future Fellowship (project number FT0991969).

## REFERENCES

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

[2] S. Baker and I. Matthews, “Lucas-kanade 20 years on: A unifying framework: Part 1: The quantity approximated, the warp update rule, and the gradient descent approximation.” *International Journal of Computer Vision*, vol. 56, no. 3, pp. 221–255, February 2004.

[3] I. Matthews and S. Baker, “Active appearance models revisited,” *International Journal of Computer Vision*, vol. 60, no. 1, pp. 135 – 164, November 2004.

[4] S. Baker and I. Matthews, “Equivalence and efficiency of image alignment algorithms,” in *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, 2001.

[5] E. Land and J. McCann, “Lightness and retinex theory,” *Journal of Optical Society of America*, vol. 61, 1971.

[6] C. Wiskott, J. M. Fellous, N. Krüger, and C. von der Malsburg, “Face recognition by elastic bunch graph matching,” *IEEE Trans. PAMI*, vol. 19, no. 7, pp. 775–779, July 1997.

[7] C. Liu and H. Wechsler, “Gabor feature based classification using the enhanced Fisher linear discriminant model for face recognition,” *IEEE Trans. Image Processing*, vol. 11, no. 4, pp. 467–476, 2002.

[8] 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*, vol. 2, pp. 568–573, June 2005.

[9] Z. Li, D. Lin, and X. Tang, “Nonparametric discriminant analysis for face recognition,” *IEEE Trans. PAMI*, vol. 31, no. 4, pp. 755–761, April 2009.

[10] A. B. Ashraf, S. Lucey, and T. Chen, “Fast image alignment in the fourier domain,” *International Conference in Computer Vision and Pattern Recognition (CVPR)*, pp. 2480–2487, 2010.

[11] R. Navarathna, S. Lucey, and S. Sridharan, “Fourier active appearance model,” *International Conference in Computer Vision (ICCV2011)*, November 2011.

[12] A. V. Oppenheim and A. S. Willsky, *Signals & Systems*, 2nd ed. Prentice Hall, 1996.

[13] G. D. Hager and P. N. Belhumeur, “Efficient region tracking with parametric models of geometry and illumination,” *IEEE Trans. PAMI*, vol. 20, pp. 1025–1039, 1998.

[14] P. N. Belhumeur and D. J. Kriegman, “What is the set of images of an object under all possible lighting conditions?” in *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, 1996.

- [15] S. Avidan, "Support vector tracking," *IEEE Trans. PAMI*, vol. 26, no. 8, pp. 1064–1072, August 2004.
- [16] S. Lucey, "Enforcing non-positive weights for stable support vector tracking," in *IEEE International Conference on Computer Vision and Pattern Recognition*, June 2008.
- [17] X. Liu, "Generic face alignment using boosted appearance model," *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–8, 2007.
- [18] M. H. Nguyen and F. De la Torre, "Local minima free parameterized appearance models," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2008.
- [19] M. Black and A. Jepson, "Eigen-tracking: Robust matching and tracking of articulated objects using a view-based representation," *International Journal of Computer Vision*, vol. 36, no. 2, pp. 101–130, 1998.
- [20] B. Theobald, "Evaluating error functions for robust active appearance models," in *IEEE International Conference on Automatic Face and Gesture Recognition*, 2006, pp. 149–154.
- [21] G. D. Evangelidis and E. Z. Psarakis, "Parametric image alignment using enhanced correlation coefficient maximization," in *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, vol. 30, no. 10, 2008, pp. 1858–1865.
- [22] N. Dowson and R. Bowden, "Mutual information for Lucas-Kanade tracking (MILK): An inverse compositional formulation," *IEEE Trans. PAMI*, vol. 30, no. 1, January 2008.
- [23] G. Tzimiropoulos, S. Zafeiriou, and M. Pantic, "Robust and efficient parametric face alignment," in *IEEE International Conference on Computer Vision (ICCV)*, 2011.
- [24] R. Gross, I. Matthews, and S. Baker, "Generic vs. person specific active appearance models," *Image and Vision Computing*, vol. 23, no. 1, pp. 1080–1093, November 2005.
- [25] S. Baker, R. Gross, I. Matthews, and T. Ishikawa, "Lucas-kanade 20 years on: A unifying framework: Part 2," Robotics Institute, Pittsburgh, PA, Tech. Rep. CMU-RI-TR-03-01, February 2003.
- [26] T. Cootes, G. Edwards, and C. Taylor, "Active appearance models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 6, pp. 681–685, 2001.
- [27] R. Gross and V. Brajovic, "An image pre-processing algorithm for illumination invariant face recognition," in *4th International Conference on Audio-and Video Based Biometric Person Authentication (AVBPA)*, Springer-Verlag, 2003, pp. 10–18.
- [28] 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*, pp. 223–230, 2006.
- [29] R. Gross, J. S. Baker, I. Matthews, and T. Kanade, "Multi-PIE," in *IEEE International Conference on Automatic Face and Gesture Recognition*, 2008.
- [30] J. F. Cohn, "Advances in behavioral science using automated facial image analysis and synthesis," *IEEE Signal Processing Magazine*, vol. 27, no. 6, November 2010.
- [31] 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*, pp. 2489–2492, Jeju Island, Korea.



**Simon Lucey** received his PhD degree from the Queensland University of Technology, Brisbane, in 2003. He is a Science Leader and a Senior Research Scientist in the Commonwealth Science and Industrial Research Organisation (CSIRO) and a current "Futures Fellow Award" recipient from the Australian Research Council. He holds adjunct Professorial positions at the University of Queensland and, Queensland University of Technology. Prof. Lucey's research is in computer vision and machine learning and their

application to human behaviour.



**Rajitha Navarathna** Rajitha Navarathna received his Bachelor of Engineering from the University of Peradeniya, Sri Lanka in 2008. Currently he is a PhD student at the Speech, Audio, Image and Video Technologies (SAIVT) lab in Queensland University of Technology, Australia. He is conducting research in the area of audio-visual speech recognition and computer vision in an automotive environment.



**Ahmed Bilal Ashraf** Ahmed Bilal Ashraf received MS and Ph.D. degrees in electrical and computer engineering both from Carnegie Mellon University, Pittsburgh, Pennsylvania, in 2006 and 2010 respectively. He is currently a post-doctoral fellow in the computational breast imaging group at the University of Pennsylvania. His research interests lie at the interface of machine learning, computer vision, and spatio-temporal image analysis.



**Sridha Sridharan** has a BSc (Electrical Engineering) degree and obtained a MSc (Communication Engineering) degree from the University of Manchester Institute of Science and Technology (UMIST), UK and a PhD degree in the area of Signal Processing from University of New South Wales, Australia. He is currently with the Queensland University of Technology (QUT) where he is a full Professor in the School of Engineering Systems. Professor Sridharan is the Deputy

Director of the Information Security Institute and the Leader of the Research Program in Speech, Audio, Image and Video Technologies at QUT. In 1997, he was the recipient of the Award of Outstanding Academic of QUT in the area of Research and Scholarship. In 2006 he received the QUT Faculty Award for Outstanding Contribution to Research. Professor Sridharan is a Senior Member of the IEEE.