The Asymmetry of Image Registration and Its Application to Face Tracking

Göksel Dedeoğlu, *Student Member*, *IEEE*, Takeo Kanade, *Fellow*, *IEEE*, and Simon Baker

Abstract—Most image registration problems are formulated in an asymmetric fashion. Given a pair of images, one is implicitly or explicitly regarded as a *template* and warped onto the other to match as well as possible. In this paper, we focus on this seemingly arbitrary choice of the roles and reveal how it may lead to biased warp estimates in the presence of relative scaling. We present a principled way of selecting the template and explain why only the *correct asymmetric* form, with the potential inclusion of a blurring step, can yield an unbiased estimator. We validate our analysis in the domain of model-based face tracking. We show how the usual Active Appearance Model (AAM) formulation overlooks the asymmetry issue, causing the fitting accuracy to degrade quickly when the observed objects are smaller than their model. We formulate a novel, "resolution-aware fitting" (RAF) algorithm that respects the asymmetry and incorporates an explicit model of the blur caused by the camera's sensing elements into the fitting formulation. We compare the RAF algorithm against a state-of-the-art tracker across a variety of resolutions and AAM complexity levels. Experimental results show that RAF significantly improves the estimation accuracy of both shape and appearance parameters when fitting to low-resolution data. Recognizing and accounting for the asymmetry of image registration leads to tangible accuracy improvements in analyzing low-resolution imagery.

Index Terms—Image registration, resolution, estimation bias, Active Appearance Models.

1 Introduction

The task of image registration underlies many computer vision applications, such as motion estimation, tracking, model-based recognition, and change detection [7], [12], [29]. Image registration is usually tackled by first defining a geometric deformation model, and then warping one image onto another such that they become as *similar* as possible according to some criterion.

We address the following questions: Given two images to register, can we treat them equally and interchangeably? What are the conditions that make a symmetric treatment of images possible? Do these conditions impose any restrictions upon applicable algorithms? Such questions are relevant to both the *formulation*, as well as the *numerical optimization* steps of the registration task.

1.1 The Problem Formulation Step

Consider, for example, the popular "sum of normed differences" objective function [19], [33]

$$\sum_{\mathbf{y}\in\mathbf{dom}I_1}\Big[I_1(\mathbf{y})-I_2\big(\mathbf{W_{12}}(\mathbf{y})\big)\Big]^p, \tag{1}$$

where I_1 and I_2 are images, $\mathbf{y}$ is a pixel coordinate in the domain of I_1, and $\mathbf{W_{12}}$ is the geometric mapping from the coordinate frame of I_1 to that of I_2. For $p=2$, this amounts to modeling I_1's pixel intensities as i.i.d. Gaussian noise added versions of those of the warped I_2. Therefore, the warp that minimizes (1) is the Maximum-Likelihood (ML) estimate, known to be asymptotically unbiased [5].

The formulation above is asymmetric: I_2 is regarded as a *template* and is warped onto I_1. Indeed, a survey of existing methods reveals that most image registration problems are formulated in a similar way and that there is rarely any discussion as to which image ought to be the template.

The asymmetry issue of (1) has been addressed in prior work [17], [18], [21], [27], [28], where, in an attempt to remove it, the objective functions were symmetrized,[1] yielding

$$\begin{aligned}\sum_{\mathbf{y}\in\mathbf{dom}I_1}\underbrace{\Big[I_1(\mathbf{y})-I_2\big(\mathbf{W_{12}}(\mathbf{y})\big)\Big]^p}_{\text{from } I_2 \text{ onto } I_1}\\ +\sum_{\mathbf{z}\in\mathbf{dom}I_2}\underbrace{\Big[I_2(\mathbf{z})-I_1\big(\mathbf{W_{21}}(\mathbf{z})\big)\Big]^p}_{\text{from } I_1 \text{ onto } I_2}.\end{aligned} \tag{2}$$

In some cases, to further impose symmetry, an additional *consistency* term on $\mathbf{W_{12}}$ and $\mathbf{W_{21}}$ has been used, such as

$$\sum_{\mathbf{y}\in\mathbf{dom}I_1}\Big[\mathbf{y}-\mathbf{W_{21}}(\mathbf{W_{12}}(\mathbf{y}))\Big]^p+\sum_{\mathbf{z}\in\mathbf{dom}I_2}\Big[\mathbf{z}-\mathbf{W_{12}}(\mathbf{W_{21}}(\mathbf{z}))\Big]^p.$$

These past approaches essentially regarded the asymmetry as an opportunity to incorporate more data and regularization priors into the problem at hand.

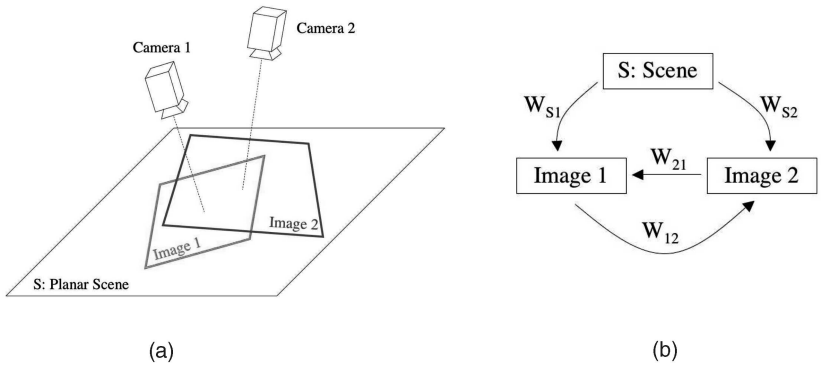
1.2 The Numerical Optimization Step

Independent of the *definition* of an objective function, its *numerical optimization* (i.e., the fitting algorithm) have also been treating the two images in an asymmetric fashion. For example, the original Lucas-Kanade algorithm [2] used a

• *The authors are with the Robotics Institute, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213.*
E-mail: {dedeoglu, tk, simonb}@cs.cmu.edu.

Manuscript received 24 Feb. 2006; revised 25 June 2006; accepted 8 Aug. 2006; published online 18 Jan. 2007.
Recommended for acceptance by P. Fua.
For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number TPAMI-0181-0206.
Digital Object Identifier no. 10.1109/TPAMI.2007.1054.

1. Reexpressing (1) in the domain of I_2 would introduce the Jacobian $|J(\mathbf{W_{12}})|$ as a weighting term. However, the symmetrized form is not necessarily limited to the original noise model. It may instead combine two noise models.



(a) (b)

Fig. 1. (a) A planar scene is observed by pinhole cameras. (b) Under central projection, scene-to-image and image-to-image transformations are homographies.

Taylor expansion of the warp around its current estimate, yielding

$$\sum_{\mathbf{y}\in\mathbf{dom}I_1}\Big[I_1(\mathbf{y})-I_2\Big(\big(\mathbf{W_{12}}+\Delta\mathbf{W_{12}}\big)(\mathbf{y})\Big)\Big]^p,$$

and iteratively solved for the warp updates $\Delta\mathbf{W_{12}}$. Observe that only image I_2 is warped in this scheme. In contrast, Baker and Matthews [31] proposed an inverse compositional algorithm that performs the expansion on I_1 and minimizes

$$\sum_{\mathbf{y}\in\mathbf{dom}I_1}\Big[I_1\big(\Delta\mathbf{W_{21}}(\mathbf{y})\big)-I_2\big(\mathbf{W_{12}}(\mathbf{y})\big)\Big]^p$$

with respect to $\Delta\mathbf{W_{21}}$, resulting in higher efficiency. Note that the inverse compositional formulation warps both images simultaneously, albeit to different degrees.

1.3 Synopsis

In this paper, we reveal a fundamental problem overlooked in the past work in both formulation and in optimization for symmetrical use of images and reveal how it may lead to biased warp estimates in the presence of relative scaling. We present a principled way of selecting for the template and explain why only the *correct* asymmetric form, with the potential inclusion of a blurring step, can yield unbiased estimators.

In Section 2, we focus on the choice of the template. In a simplified image registration scenario, we investigate when and why the selection of the template should be of concern. We establish that a relative scaling between images imposes a particular asymmetric form for obtaining unbiased estimates of the underlying geometric transformation.

In Section 3, we validate our analysis in the domain of model-based face tracking. In reviewing the popular Active Appearance Model (AAM) formulation, we observe how it also overlooks the asymmetry issue. As predicted, the AAM fitting accuracy is shown to degrade quickly when the observed faces become blurred and are smaller than the model. We then formulate a novel, "resolution-aware fitting" (RAF) algorithm that respects the asymmetry, and incorporates an explicit model of the blur caused by the camera's sensing elements. We experimentally compare the new algorithm against a state-of-the-art tracker across a variety of resolution and AAM complexity levels.

In Section 4, we discuss some practical and algorithmic consequences of the asymmetry and identify directions for future investigation.

2 Analysis of the Image Registration Problem

2.1 A Simplified Scenario

Without loss of generality, let us consider the simplified scenario shown in Fig. 1a, in which a planar scene S is observed by two pinhole cameras which capture continuous images. Under the central projection model, scene-to-image and image-to-image coordinate transformations will be homographies [23]. Note that this class of geometric transformation will account for observed images *exactly*. In order to avoid complications arising from noncorresponding image points, we assume that both images have infinite extent and are free of occlusion.

As shown in Fig. 1b, the domains of the scene radiance S, image I_1, and image I_2 are related by homographies. $\mathbf{W_{S1}}$ and $\mathbf{W_{S2}}$ denote transformations which take scene coordinates and compute their corresponding image point locations in I_1 and I_2, respectively. $\mathbf{W_{12}}$ denotes the transformation from I_1 to I_2, and $\mathbf{W_{21}}$ from I_2 to I_1. To render the problem as well-posed as possible, all transforms are assumed to be invertible, i.e., $\mathbf{W_{12}}=\mathbf{W_{21}^{-1}}$. Thus, the image registration task is to estimate the homography $\mathbf{W_{12}}$ (or $\mathbf{W_{21}}$) between I_1's and I_2's coordinate frames based on image intensity measurements.

We will use two equivalent notations to express the fact that one image is a geometrically transformed version of another. The first one is $I_1(\mathbf{y})=I_2(\mathbf{W_{12}}(\mathbf{y}))$. Using point coordinates, this notation indicates where a particular image point maps onto the other image and states how those image intensities relate to each other. Alternatively, we will use $I_1=warp(I_2;\mathbf{W_{21}})$. This notation refers to an entire domain's transformation. It states that I_1 is the image obtained by transforming every point in the domain of I_2 by $\mathbf{W_{21}}$; note the use of $\mathbf{W_{21}}$ here instead of $\mathbf{W_{12}}$, since the transformed points are in I_2.

2.2 Theoretical Case: Ideal Camera and Known Scene

We start our discussion with an idealized case. Suppose that we have full knowledge of the underlying scene radiance function S, and both cameras are ideal; their lenses precisely focus incoming light rays parallel to the optical axis onto the camera's image plane and their photo-receptive fields are continuous (i.e., they have infinite resolution). We model the intensity at an image point as a noisy (i.i.d., additive

Gaussian) observation of the corresponding scene point's radiance,

$$I_1(\mathbf{y}) = S\big(\mathbf{W_{1S}}(\mathbf{y})\big) + \epsilon(\mathbf{y}) \qquad \forall \mathbf{y} \in \mathbf{dom} I_1, \tag{3}$$

$$I_2(\mathbf{z}) = S\big(\mathbf{W_{2S}}(\mathbf{z})\big) + \epsilon(\mathbf{z}) \qquad \forall \mathbf{z} \in \mathbf{dom} I_2, \tag{4}$$

where $\mathbf{y}$ and $\mathbf{z}$ are points in the domains of I_1 and I_2, respectively. We denote image-to-scene warps (homography) by $\mathbf{W_{1S}}$ and $\mathbf{W_{2S}}$ (Fig. 1b). Using the alternative notation, (3) and (4) can be also expressed as

$$I_1(\mathbf{y}) = warp(S; \mathbf{W_{S1}})(\mathbf{y}) + \epsilon(\mathbf{y}) \qquad \forall \mathbf{y} \in \mathbf{dom} I_1, \tag{5}$$

$$I_2(\mathbf{z}) = warp(S; \mathbf{W_{S2}})(\mathbf{z}) + \epsilon(\mathbf{z}) \qquad \forall \mathbf{z} \in \mathbf{dom} I_2. \tag{6}$$

In the following, we present three equivalent methods which would compute the ML estimate of $\mathbf{W_{12}}$. Given our assumptions at this moment, these algorithms are rather trivial. Nevertheless, they will be minimally affected while the assumptions are relaxed later in the paper, allowing us to highlight their applicability to different situations.

A1. Generative Algorithm
Step 1: Find the ML parameters for scene-to-image warps $\mathbf{W_{S1}}$ and $\mathbf{W_{S2}}$:

$$\hat{\mathbf{W}}_{\mathbf{S1}} = \arg\min_{\mathbf{W_{S1}}} \int_{\mathbf{y} \in \mathbf{dom} I_1} \Big[I_1(\mathbf{y}) - warp(S; \mathbf{W_{S1}}(\mathbf{y}) \Big]^2 \mathbf{dy}. \tag{7}$$

$$\hat{\mathbf{W}}_{\mathbf{S2}} = \arg\min_{\mathbf{W_{S2}}} \int_{\mathbf{z} \in \mathbf{dom} I_2} \Big[I_2(\mathbf{z}) - warp(S; \mathbf{W_{S2}}(\mathbf{z}) \Big]^2 \mathbf{dz}. \tag{8}$$

Step 2: Compose them to obtain the ML estimate of the relative warp $\mathbf{W_{12}}$:

$$\hat{\mathbf{W}}_{\mathbf{12}} = \hat{\mathbf{W}}_{\mathbf{1S}} \circ \hat{\mathbf{W}}_{\mathbf{S2}} = (\hat{\mathbf{W}}_{\mathbf{S1}})^{-1} \circ \hat{\mathbf{W}}_{\mathbf{S2}}.$$

B1. Forward Algorithm
Step 1: Find $\hat{\mathbf{W}}_{\mathbf{S1}}$ and $\hat{\mathbf{W}}_{\mathbf{S2}}$ by (7) and (8).
Step 2: Based on the scene function S and ML estimates $\hat{\mathbf{W}}_{\mathbf{S1}}$ and $\hat{\mathbf{W}}_{\mathbf{S2}}$, set up a direct estimation problem for the relative warp $\mathbf{W_{12}}$:

$$\hat{\mathbf{W}}_{\mathbf{12}} = \arg\min_{\mathbf{W_{12}}} \int_{\mathbf{z} \in \mathbf{dom} I_2} \Big[\underbrace{warp(S; \hat{\mathbf{W}}_{\mathbf{S2}})}_{\hat{I}_2}(\mathbf{z}) - warp\big(\underbrace{warp(S; \hat{\mathbf{W}}_{\mathbf{S1}})}_{\hat{I}_1}; \mathbf{W_{12}} \big)(\mathbf{z}) \Big]^2 \mathbf{dz}. \tag{9}$$

By computing $\hat{I}_1 = warp(S; \hat{\mathbf{W}}_{\mathbf{S1}})$ and $\hat{I}_2 = warp(S; \hat{\mathbf{W}}_{\mathbf{S2}})$, this method essentially *simulates* the formation of ML images of I_1 and I_2. In other words, the registration problem is posed in terms of ML images:

$$\hat{\mathbf{W}}_{\mathbf{12}} = \arg\min_{\mathbf{W_{12}}} \int_{\mathbf{z} \in \mathbf{dom} I_2} \Big[\hat{I}_2(\mathbf{z}) - warp(\hat{I}_1; \mathbf{W_{12}})(\mathbf{z}) \Big]^2 \mathbf{dz}. \tag{10}$$

Note the similarity between (10) and (1): They are both asymmetric and warp only one of the images. Indeed, one can use images I_1 and I_2 as plug-in estimates of $\hat{I}_1$ and $\hat{I}_2$ and directly estimate $\mathbf{W_{12}}$. This seems to be exactly the idea behind commonly used objective functions such as (1).

C1. Backward Algorithm
Step 1: Find $\hat{\mathbf{W}}_{\mathbf{S1}}$ and $\hat{\mathbf{W}}_{\mathbf{S2}}$ by (7) and (8).
Step 2: Just as in the *forward* algorithm B1, set up a new warp estimation problem. This time, however, solve for the warp in the opposite direction by warping the other ML image:

$$\hat{\mathbf{W}}_{\mathbf{21}} = \arg\min_{\mathbf{W_{21}}} \int_{\mathbf{y} \in \mathbf{dom} I_1} \Big[\underbrace{warp(S; \hat{\mathbf{W}}_{\mathbf{S1}})}_{\hat{I}_1}(\mathbf{y}) - warp\big(\underbrace{warp(S; \hat{\mathbf{W}}_{\mathbf{S2}})}_{\hat{I}_2}; \mathbf{W_{21}} \big)(\mathbf{y}) \Big]^2 \mathbf{dy}. \tag{11}$$

We have intentionally defined both algorithms to be asymmetric: The *forward* Algorithm B1 warps $\hat{I}_1$ onto $\hat{I}_2$, and the *backward* Algorithm C1 does the opposite. Using this setup, we can investigate whether there is a fundamental difference between the two. In the Appendix, we show that the ML warp estimates of all three algorithms would be the same for similarity transforms.

Choosing an Algorithm. We have shown that, under the assumptions about the camera and those in the Appendix on the warp, all three formulations of the problem generate the same ML estimate of the warp. The *generative* one requires the knowledge of the scene, but the asymmetric *forward* and *backward* methods do not because their Step 1 can be skipped and I_1 and I_2 used instead.

In the next section, we weaken our assumptions and show that the equivalence of the formulations above is no longer true. While the generative formulation can still give us an ML estimate, the asymmetric ones do not (unless modified appropriately).

2.3 Practical Case: Real Camera and Unknown Scene

A real camera has blur effects. The response of a camera to an ideal point light source is characterized by its point spread function (PSF). This means that the scene irradiance will be subject to a convolution with the PSF. For convenience, we still assume the images to be continuous. Instead of (5) and (6), we have

$$I_1(\mathbf{y}) = B\big(warp(S; \mathbf{W_{S1}})\big)(\mathbf{y}) + \epsilon_1(\mathbf{y}) \quad \forall \mathbf{y} \in \mathbf{dom} I_1, \tag{12}$$

$$I_2(\mathbf{z}) = B\big(warp(S; \mathbf{W_{S2}})\big)(\mathbf{z}) + \epsilon_2(\mathbf{z}) \quad \forall \mathbf{z} \in \mathbf{dom} I_2, \tag{13}$$

where the blur operator $B(\cdot)$ indicates a convolution with the PSF:

$$B\big(S\big)(\mathbf{x}) = \int_{\mathbf{w} \in \mathbf{dom} S} S(\mathbf{w}) PSF(\mathbf{w} - \mathbf{x}) \mathbf{dw}.$$

Due to imperfect lenses and density constraints on photoreceptive sensing elements, the PSF of a real camera is not a delta function [1]. In fact, the PSF is closely related to measurement noise characteristics. In order to operate at prescribed frame rates and signal-to-noise ratio levels, CCD cameras accumulate photon counts over a finite spatial extent, a procedure called *binning*. The blur model must not only account for realistic lens optics, but also capture those binning operations which take place at the sensing element level.

Also, in real situations, we do not know the scene radiance S. Assuming a blurry camera and unknown scene, let us discuss the three algorithms corresponding to those considered in Section 2.2 for the ideal case.

A2. Generative Algorithm

Step 1:

$$\hat{\mathbf{W}}_{\mathbf{S1}} = \arg\min_{\mathbf{W_{1S}}} \int\limits_{\mathbf{y}\in\mathbf{dom}I_1} \Big[I_1(\mathbf{y}) - B\big(warp(S;\mathbf{W_{S1}})\big)(\mathbf{y})\Big]^2 \mathbf{dy}, \tag{14}$$

$$\hat{\mathbf{W}}_{\mathbf{S2}} = \arg\min_{\mathbf{W_{2S}}} \int\limits_{\mathbf{z}\in\mathbf{dom}I_2} \Big[I_2(\mathbf{z}) - B\big(warp(S;\mathbf{W_{S2}})\big)(\mathbf{z})\Big]^2 \mathbf{dz}. \tag{15}$$

Step 2: $\hat{\mathbf{W}}_{\mathbf{12}} = \hat{\mathbf{W}}_{\mathbf{1S}} \circ \hat{\mathbf{W}}_{\mathbf{S2}} = (\hat{\mathbf{W}}_{\mathbf{S1}})^{-1} \circ \hat{\mathbf{W}}_{\mathbf{S2}}$.

Since we do not know the scene radiance S, we need to estimate it jointly with the warps. This approach was proposed in the past as part of a superresolution problem by [10].

Although theoretically sound and elegant, the *generative* algorithm is rarely used in registering images. Instead, *forward* or *backward* algorithms that perform image-to-image comparisons as in (1) are used, presuming their equivalence. In the presence of camera blur, however, this turns out to be incorrect.

B2. Forward Algorithm

The corresponding algorithm to that in B1 is

Step 1: Find $\hat{\mathbf{W}}_{\mathbf{S1}}$ and $\hat{\mathbf{W}}_{\mathbf{S2}}$ by (14) and (15).

Step 2:

$$\hat{\mathbf{W}}_{\mathbf{12}} = \arg\min_{\mathbf{W_{12}}} \int\limits_{\mathbf{z}\in\mathbf{dom}I_2} \Big[\underbrace{B\big(warp(S;\hat{\mathbf{W}}_{\mathbf{S2}})\big)}_{\hat{I}_2}(\mathbf{z}) - \underbrace{B\Big(warp\big(warp(S;\hat{\mathbf{W}}_{\mathbf{S1}});\mathbf{W_{12}}\big)\Big)}_{T}(\mathbf{z})\Big]^2 \mathbf{dz}. \tag{16}$$

This algorithm could generate the ML estimate only if the scene S was known. The immediate question is whether we can follow the same steps as before and use the images I_1 and I_2 in place of the warped scene. In the presence of blur, this turns out to be not always possible.

Note that the observed image I_2 is the ML estimate for

$$\hat{I}_2 = B\big(warp(S;\hat{\mathbf{W}}_{\mathbf{S2}})\big).$$

Suppose we denote by T the following "imaging" function

$$T = B\Big(warp\big(warp(S;\hat{\mathbf{W}}_{\mathbf{S1}});\mathbf{W_{12}}\big)\Big).$$

Then, the image registration problem of (16) becomes

$$\hat{\mathbf{W}}_{\mathbf{12}} = \arg\min_{\mathbf{W_{12}}} \int\limits_{\mathbf{z}\in\mathbf{dom}I_2} \Big[I_2(\mathbf{z}) - T(\mathbf{z})\Big]^2 \mathbf{dz}. \tag{17}$$

Since T is still a function of the unknown S, it cannot be readily computed. For the sake of argument, consider changing the order of warp and blur operators in T, and define a new imaging function

$$T' = warp\Big(\underbrace{B\big(warp(S;\hat{\mathbf{W}}_{\mathbf{S1}})\big)}_{\hat{I}_1};\mathbf{W_{12}}\Big).$$

The observed image I_1 is the ML estimate for $\hat{I}_1 = B(warp(S;\hat{\mathbf{W}}_{\mathbf{S1}}))$ and, therefore, $T' = warp(I_1;\mathbf{W_{12}})$. That is, if we replace T by T' in (17), we would arrive at the commonly used form (1) of objective function in image registration (for $p = 2$):

$$\hat{\mathbf{W}}'_{\mathbf{12}} = \arg\min_{\mathbf{W_{12}}} \int\limits_{\mathbf{z}\in\mathbf{dom}I_2} \Big[I_2(\mathbf{z}) - warp(I_1;\mathbf{W_{12}})(\mathbf{z})\Big]^2 \mathbf{dz}. \tag{18}$$

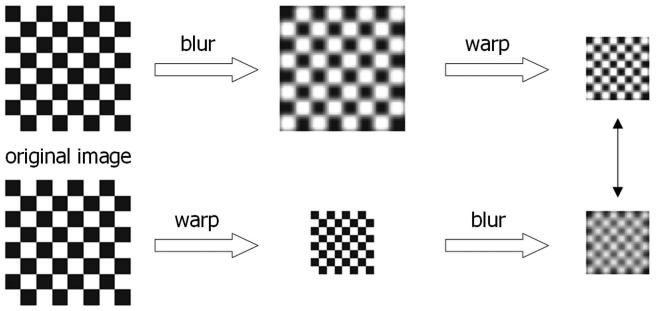


Fig. 2. The order of blurring and geometric warp operations is important: In this example, we used the same Gaussian blur kernel ($\sigma = 2$ pixels) before (top row) or after (bottom row) geometric scaling by a factor of 1/2. Resulting images, shown on the right, differ from each other.

However, the warp and blur operations do not commute in general. Fig. 2 illustrates this fact with a simple example. We therefore have $T \neq T'$, resulting in $\hat{\mathbf{W}}'_{12} \neq \hat{\mathbf{W}}_{12}$. Since $\hat{\mathbf{W}}'_{12}$ does not coincide with the ML solution $\hat{\mathbf{W}}_{12}$, it will be a *biased* estimator.

Compensating for the Bias. While $\hat{\mathbf{W}}'_{12}$ is biased, there exist conditions under which T' can help us compute the unbiased estimate $\hat{\mathbf{W}}_{12}$. To reveal when this would be possible, we express the blur operators in T and T' explicitly as convolution integrals. For notational conciseness, let us define $S' = warp(S;\hat{\mathbf{W}}_{\mathbf{S1}})$.

$$\begin{aligned} T(\mathbf{x}) &= B\Big(warp\big(warp(S;\hat{\mathbf{W}}_{\mathbf{S1}});\mathbf{W_{12}}\big)\Big)(\mathbf{x}) \\ &= B\big(warp(S';\mathbf{W_{12}})\big)(\mathbf{x}) \\ &= \int\limits_{\mathbf{w}\in\mathbf{dom}\ warp(S';\mathbf{W_{12}})} warp(S';\mathbf{W_{12}})(\mathbf{w})PSF\big(\mathbf{w}-\mathbf{x}\big)\mathbf{dw}. \end{aligned} \tag{19}$$

On the other hand,

$$\begin{aligned} T'(\mathbf{x}) &= warp\Big(B\big(warp(S;\hat{\mathbf{W}}_{\mathbf{S1}})\big);\mathbf{W_{12}}\Big)(\mathbf{x}) \\ &= warp\big(B(S');\mathbf{W_{12}}\big)(\mathbf{x}) \\ &= B\big(S'\big)\big(\mathbf{W}_{\mathbf{12}}^{-1}(\mathbf{x})\big) \\ &= \int\limits_{\mathbf{v}\in\mathbf{dom}S'} S'(\mathbf{v})PSF\big(\mathbf{v}-\mathbf{W}_{\mathbf{12}}^{-1}(\mathbf{x})\big)\mathbf{dv}. \end{aligned}$$

To rewrite the integral above in the domain of $warp(S';\mathbf{W_{12}})$, we define $\mathbf{w} = \mathbf{W_{12}}(\mathbf{v})$. As $\mathrm{dv} = |J(\mathbf{W}_{\mathbf{12}}^{-1})|\mathrm{dw}$, changing the variable of integration of $\mathbf{v}$ to $\mathbf{w}$ will yield

$$T'(\mathbf{x}) = \int_{\mathbf{w}\in \mathbf{dom}\ warp(S';\mathbf{W_{12}})} warp(S';\mathbf{W_{12}})(\mathbf{w}) \\ PSF\big(\underbrace{\mathbf{W_{12}}^{-1}(\mathbf{w}) - \mathbf{W_{12}}^{-1}(\mathbf{x})}_{!}\big)\,\big|J(\mathbf{W_{12}}^{-1})\big|\mathbf{dw}. \tag{20}$$

Intuition for the Restricted Case of a Similarity Transform. Observe that the difference between T in (19) and T' in (20) is due to the transformation of PSF's argument in (20). Before discussing general properties of this difference and providing a concrete method for its elimination, we can develop an intuition for a restricted case.

Let us consider $\mathbf{W_{12}}$ to be a similarity transformation, which can be parameterized using scale s, rotation θ, and translation (t_x, t_y) variables. The argument of the PSF in (20) is then

$$\begin{aligned}
&\mathbf{W_{12}}^{-1}(\mathbf{w}) - \mathbf{W_{12}}^{-1}(\mathbf{x}) \\
&= \left(\begin{bmatrix} scos\theta & -ssin\theta \\ ssin\theta & scos\theta \end{bmatrix}^{-1}\begin{bmatrix} w_x \\ w_y \end{bmatrix} - \begin{bmatrix} t_x \\ t_y \end{bmatrix}\right) \\
&\quad - \left(\begin{bmatrix} scos\theta & -ssin\theta \\ ssin\theta & scos\theta \end{bmatrix}^{-1}\begin{bmatrix} x_x \\ x_y \end{bmatrix} - \begin{bmatrix} t_x \\ t_y \end{bmatrix}\right) \\
&= \begin{bmatrix} \frac{cos\theta}{s} & \frac{sin\theta}{s} \\ \frac{-sin\theta}{s} & \frac{cos\theta}{s} \end{bmatrix}\begin{bmatrix} w_x \\ w_y \end{bmatrix} - \begin{bmatrix} \frac{cos\theta}{s} & \frac{sin\theta}{s} \\ \frac{-sin\theta}{s} & \frac{cos\theta}{s} \end{bmatrix}\begin{bmatrix} x_x \\ x_y \end{bmatrix} \\
&= \begin{bmatrix} \frac{cos\theta}{s} & \frac{sin\theta}{s} \\ \frac{-sin\theta}{s} & \frac{cos\theta}{s} \end{bmatrix}\begin{bmatrix} w_x - x_x \\ w_y - x_y \end{bmatrix} \\
&= \mathbf{W}'(\mathbf{w} - \mathbf{x}),
\end{aligned}$$

where $\mathbf{W}'$ is a similarity transform with scale $\frac{1}{s}$, rotation θ, and zero translation. Furthermore, if the camera's PSF is rotation-invariant (i.e., isotropic),

$$PSF\big(\mathbf{W}'(\mathbf{w}-\mathbf{x})\big) = PSF\Big(\frac{\mathbf{w}-\mathbf{x}}{s}\Big).$$

In summary, when $\mathbf{W_{12}}$ is limited to similarity transforms and the PSF is isotropic, (20) becomes

$$T'(\mathbf{x}) = \int_{\mathbf{w}\in \mathbf{dom}\ warp(S';\mathbf{W_{12}})} warp(S';\mathbf{W_{12}})(\mathbf{w})\, PSF\Big(\frac{\mathbf{w}-\mathbf{x}}{s}\Big)\big|J\big(\mathbf{W_{12}}^{-1}\big)\big|\mathbf{dw}. \tag{21}$$

A comparison of (21) with (19) reveals how the imaging functions T' and T relate to each other. Although they are both obtained by blurring $warp(S';\mathbf{W_{12}})$, the actual blur kernels are different. Imagine that T has the blur kernel $PSF(\cdot)$, shown in Fig. 3b. Since the blur kernel of T' is $PSF(\frac{\cdot}{s})$, it will have a dilated or compressed shape. For $0 < s < 1$, the kernel gets compressed (Fig. 3a), resulting in a T' less blurry than T. For $s = 1$, we have equality between T and T'. Finally, for $s > 1$, the effective blur kernel becomes wider (Fig. 3c), causing T' to be even more blurred than T.

The analysis above provides the conditions under which T' can be used in emulating T, and the *forward* algorithm B2 still work, even if the scene function S is unknown:

- For $s = 1$, T' can readily replace T.
- For $0 < s < 1$, we may blur T' further to make up for the difference in blur kernels $PSF(\cdot)$ and $PSF(\frac{\cdot}{s})$. Only after this blur compensation is made would the minimizer of (18) correspond to the unbiased ML estimate.
- For $s > 1$, the wider blurring kernel produces an overblurred T', and emulating T then turns out to be a deblurring problem: This is, in general, an ill-posed inverse problem and difficult to solve [11].

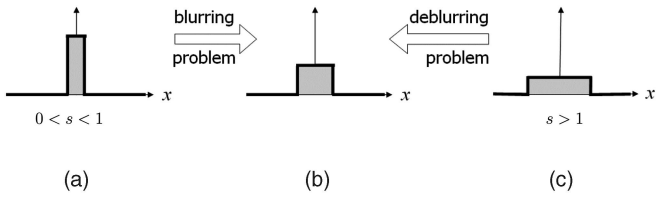


Fig. 3. Given I_1, we can compute $T' = \mathbf{W_{12}}(I_1)$. However, to find the ML warp estimates, we need to evaluate T, which originally results from a (b) convolution operation with the PSF of the camera. Depending on the value of s, estimating T from T' turns out to be a blurring ($0 < s < 1$) or deblurring ($s > 1$) problem.

Note that the quantities T and T' in (19) and (21) were derived for the *forward* algorithm. By definition, when the *forward* algorithm scales up (i.e., $s > 1$), the *backward* algorithm scales down ($0 < \frac{1}{s} < 1$). Therefore, in situations where $s > 1$, the deblurring problem can be avoided by simply switching to the algorithm which solves for the warp in the opposite direction. Hence, for obtaining an unbiased estimate of the warp between two images, there is a *natural choice* between the *forward* and *backward* algorithms: One should pick the direction of warp such that, after necessary blurring, it scales one image down onto the other, i.e., the higher resolution image should be warped onto the lower resolution image.[2]

More General Cases. For more complex warps than similarity, the blur varies spatially and the analysis above does not apply. The inequality between T and T' is still due to the difference between the PSF's arguments in (19) and (20), but the analysis becomes harder. Probably the general solution is to use the *generative* algorithm and explicitly recover S.

In specific cases, however, it may be possible to derive an algorithm, but this has to be done on a case-by-case basis. In the next section, we do this for the specific case of piecewise affine warps used in a face model.

C2. Backward Algorithm

The above analysis also applies to the *backward* algorithm.

3 Validating the Asymmetry and Quantifying its Effects

We have found out that the potential asymmetry between *forward* and *backward* registration algorithms is due to the difference in their effective blur kernel size. Our analysis indicates that this difference becomes more pronounced as the relative magnification factor between images becomes larger. Therefore, in order to obtain an unbiased warp estimate, one must start with the higher resolution image

2. Could one still blur the high-resolution image, but warp the low-resolution image onto the higher resolution one instead? It is hard to tell which optimization criterion this approach would be minimizing and whether its solution would correspond to the ML estimate of the warp.

and then warp it onto the lower resolution one, incorporating a model of the blur-formation process in the fitting criterion. On the other hand, if the scaling-induced blur effect is ignored, or the lower resolution image is warped (and interpolated) onto the higher resolution one, one should expect the warp estimates to be biased. In this section, we demonstrate and measure this bias in realistic cases.

One cannot quantify blur effects independently from the image content: While blurring (i.e., low-pass filtering) visually rich and detailed images would make a significant difference, it would barely alter already smooth images. This consideration led us to consider a particular class of images, namely, those of human faces. Accurate registration algorithms are crucial in this domain because they determine the performance of various tracking, recognition, and biometric verification systems. We chose to verify our bias predictions and present a specific solution in a model-based face tracking application.

Active Appearance Models (AAM) are compact parametric representations of the shape and appearance of objects [13], [25] and have been most popular in tracking human faces. Such models are typically built at nominal image resolutions, where the landmarks describing the shape of a face (such as eyebrows and lips) are localized and manually marked over a set of training images. Once an AAM has been learned, it is fit to a new image for interpretation. A *fitting* algorithm recovers those parameters which *best* explain a given image. This nonlinear optimization task is one of the image registration problems covered in Section 2, except that an AAM has many more parameters to fit: The warp is more complex, a piecewise affine warp defined by a collection of nonrigid shape modes. An AAM also has a linear appearance model that has to be solved for.

In this section, we consider cases in which observed faces are lower resolution than the model and focus on AAM fitting accuracy metrics. We examine the traditional AAM fitting formulation in the light of our bias analysis and reveal how it overlooks the asymmetry issue. As predicted, the fitting accuracy is shown to degrade quickly in lower resolutions. We then propose a new ML algorithm which respects the asymmetry and incorporates a model of the camera blur, leading to significantly more accurate fitting results.

3.1 Active Appearance Models

An AAM [13], [25] consists of two models, namely, the *shape* and *appearance* of an object. Each of these is a linear Principal Components model learned from training data. The shape of an AAM is defined by a set of 2D landmark locations

$$\mathbf{s} = (x_1, y_1, x_2, y_2, \ldots, x_v, y_v)^{\mathrm{T}}. \tag{22}$$

The shape model, parametrized with $\mathbf{p} = (p_1, p_2, \ldots, p_n)$, expresses any shape as a linear combination of basis shapes added onto a base shape:

$$\mathbf{s}(\mathbf{p}) = \mathbf{s}_0 + \sum_{i=1}^{n} p_i \mathbf{s}_i. \tag{23}$$

An AAM is defined in the coordinate system of the object being modeled. To express object instances in arbitrary poses, a global transform is needed. Following [32], we define four special shape bases to account for similarity transforms (scale, rotation, and two translations) and compose them with the shape model. We denote the combined geometric deformation by $\mathbf{W}(\mathbf{x}; \mathbf{p})$, where $\mathbf{x}$ is a model point coordinate being mapped onto an image coordinate.

The appearance model consists of the mean and basis images. The basis images are shape-normalized, i.e., they are defined within the base shape $\mathbf{s}_0$. The appearance model is linear and parametrized with $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_m)$ as

$$A(\mathbf{x}; \boldsymbol{\lambda}) = A_0(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i A_i(\mathbf{x}) \qquad \forall\, \mathbf{x} \in \mathbf{dom}\ \mathbf{s}_0, \tag{24}$$

where $\mathbf{x}$ is a pixel coordinate in $\mathbf{s}_0$. The appearance basis images are usually defined at reasonable resolution levels such as 100×100 pixels for face models.

In this paper, we consider the simpler case of *independent* AAMs [32], where the statistical dependence between the shape and appearance is ignored. While such couplings have been exploited in prior work [13], [25], their advantages remain orthogonal to our discussion.

3.2 Traditional Fitting Formulation

Given a set of AAM parameters, the linear generative equations (23) and (24) can uniquely synthesize an object instance [32]. Image analysis deals with the inverse of this process. It aims to recover those AAM parameters which *best* explain a given image. For this end, one needs to define a similarity metric to quantify what constitutes a good match, and a *fitting* algorithm for computing the parameter values which optimize the similarity metric. The fitting criterion also specifies the direction of warp, i.e., whether the template $A(\boldsymbol{\lambda})$ ought to be warped onto the observed image or vice versa.

In the original AAM work by Cootes et al. [13], [14], [25], as well as its computationally efficient reformulation by Matthews and Baker [32], the fitting criterion was the sum of squared intensity differences between the synthesized model template and the *warped input image* I:

$$\sum_{\mathbf{x} \in \mathbf{dom}\ \mathbf{s}_0} \Big[I\big(\mathbf{W}(\mathbf{x}; \mathbf{p})\big) - A(\mathbf{x}; \lambda) \Big]^2. \tag{25}$$

Since this objective function is highly nonlinear in its parameters, iterative gradient-descent methods are usually used: In each iteration, updates $\Delta\mathbf{p}$ and $\Delta\boldsymbol{\lambda}$ are computed and added to (or composed with) current estimates of $\mathbf{p}$ and $\boldsymbol{\lambda}$, respectively. Cootes et al. [13], [14], [25] assumed a constant, linear relationship between the error image and the additive updates. They learned this mapping through regression on perturbation-based training data. Matthews and Baker [32] showed that, in general, there is no constant linear relationship between the error image and the update in the additive case, but that there is in the (inverse) compositional case. Based on this insight, and using the independence of the shape and appearance models in an independent AAM, they derived an efficient AAM fitting algorithm that runs at over 200 frames per second on typically sized AAMs.

Note that the summation in (25) is defined over $\mathbf{x}$, pixel coordinates in the shape-normalized template image $A(\boldsymbol{\lambda})$. Fig. 4 visualizes this procedure, where $\mathbf{u}$ denotes the pixel coordinates of a low-resolution input image I. Observe how the fitting criterion prescribes *first warping and interpolating* the image I and *then* comparing it against the synthesized template. The latter is normalized to shape $\mathbf{s}_0$ at the AAM's

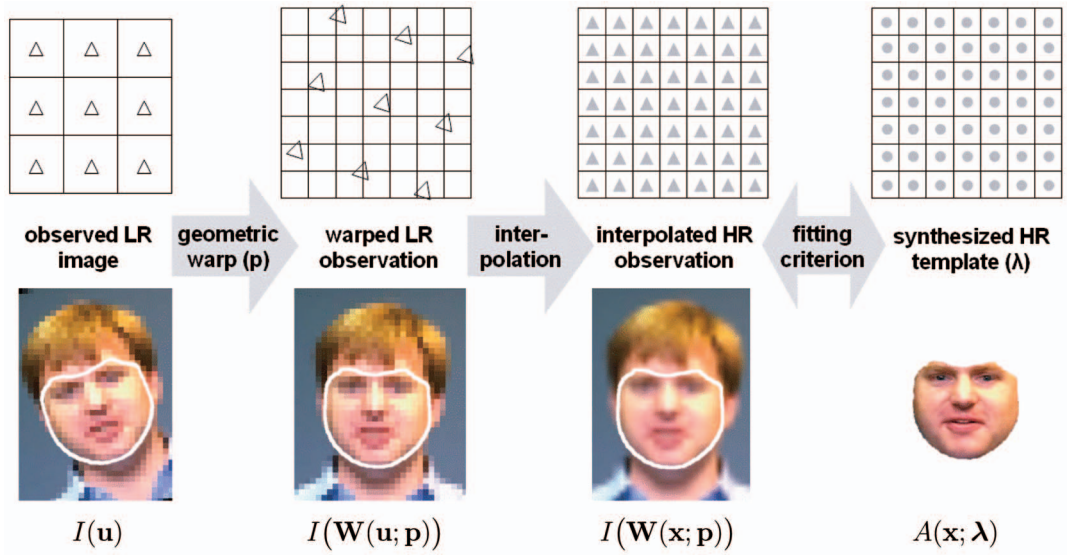


Fig. 4. Graphical representation of the traditional fitting criterion of (25). From left to right, observed images get warped, interpolated, and, finally, compared against the synthesized model instance. When the input image is low in resolution, significant interpolation is needed to warp it onto the model coordinate frame.

native resolution and remains fixed in size. Consequently, when objects appear small in comparison to the AAM, they need to be enlarged through interpolation. Recalling our asymmetry analysis, we should expect the fitting results to be increasingly biased with higher scaling factors because the fitting *criterion* itself does not respect the asymmetry of the problem at hand. Using the same gradient-descent algorithm and low-resolution images, but minimizing a more carefully designed fitting criterion, we can indeed improve the fitting accuracy.

3.3 Resolution-Aware Fitting (RAF)

3.3.1 Formulation

We propose an alternative to the fitting criterion (25). Recognizing the asymmetry of the problem, we not only change the warp direction, but also introduce a model of the blur/image formation process. From a generative point of view, this simulates the pixel-wise image formation process in a CCD camera [1]. We feed the AAM and its current parameters into a camera model and compare the outcome against the observed low-resolution image. Mathematically, the proposed fitting criterion is

$$\sum_{\mathbf{u}\in \mathrm{dom} I} \left[I(\mathbf{u}) - B\big(\mathbf{u}; A(\mathbf{W}(\mathbf{p});\boldsymbol{\lambda})\big)\right]^2, \tag{26}$$

where the summation is now over pixel coordinates $\mathbf{u}$ of the observed image I. That is, if (25) was the *forward* algorithm of Section 2, (26) is the *backward* algorithm with an additional blur operator B. This blur simulates a low-resolution image of the object, believed to be what the camera would have captured under current AAM parameters.[3] Although this formulation can accommodate arbitrary camera models and point spread functions, in this paper, we use the rectangular PSF

$$B\big(\mathbf{u}; A(\mathbf{W}(\mathbf{p});\boldsymbol{\lambda})\big) = \frac{1}{area(\mathbf{u})} \int\limits_{\mathbf{u}'\in bin(\mathbf{u})} A\big(\mathbf{W}^{-1}(\mathbf{u}';\mathbf{p});\boldsymbol{\lambda}\big)d\mathbf{u}',$$

where the continuous integral is defined over $bin(\mathbf{u})$, the sensing area of the discrete pixel $\mathbf{u}$. As illustrated in Fig. 5, the blur operator itself is independent of AAM parameters. It simply averages out those template pixel intensities which map into a low-resolution pixel's sensing area under the current warp $\mathbf{p}$. To express the integral above in the shape-normalized coordinate frame $\mathbf{s}_0$, we observe that $\mathbf{u}' = \mathbf{W}(\mathbf{x};\mathbf{p})$ and, consequently, $d\mathbf{u}' = |J(\mathbf{W}(\mathbf{p}))|d\mathbf{x}$,

$$B\big(\mathbf{u}; A(\mathbf{W}(\mathbf{p});\boldsymbol{\lambda})\big) = \frac{1}{area(\mathbf{u})} \int\limits_{\substack{\mathbf{x}\in \mathbf{dom}\, s_0 \text{ s.t.} \\ \mathbf{W}(\mathbf{x};\mathbf{p})\in bin(\mathbf{u})}} A(\mathbf{x};\boldsymbol{\lambda})\big|J\big(\mathbf{W}(\mathbf{p})\big)\big|d\mathbf{x}.$$

In practice, we implement this integration as a discrete, Jacobian-weighted sum over template pixels,

$$B\big(\mathbf{u}; A(\mathbf{W}(\mathbf{p});\boldsymbol{\lambda})\big) = \frac{1}{area(\mathbf{u})} \sum_{\substack{\mathbf{x}\in \mathbf{dom}\, s_0 \text{ s.t.} \\ \mathbf{u}-\begin{bmatrix}.5\\.5\end{bmatrix}<\mathbf{W}(\mathbf{x};\mathbf{p})<\mathbf{u}+\begin{bmatrix}.5\\.5\end{bmatrix}}} A(\mathbf{x};\boldsymbol{\lambda})\big|J\big(\mathbf{W}(\mathbf{p})\big)\big|. \tag{27}$$

Observe that our formulation avoids interpolating low-resolution data and models the object appearance, geometric deformation, and the image formation processes simultaneously.

3.3.2 RAF Algorithm

We now present a Gauss-Newton gradient-descent algorithm for the minimization of the fitting criterion (26) with respect to $\mathbf{p}$ and $\boldsymbol{\lambda}$. This algorithm gives up the computational efficiency of [32] in exchange for a more accurate/unbiased estimate of the parameters. Until convergence, updates $\Delta\mathbf{p}$ and $\Delta\boldsymbol{\lambda}$ are iteratively computed and added to the current estimates. The derivation below closely follows that of the *simultaneous* algorithm in [30]. Expressing $A(\boldsymbol{\lambda})$

3. If the camera is expected to have aliasing, our method should simulate that as well.

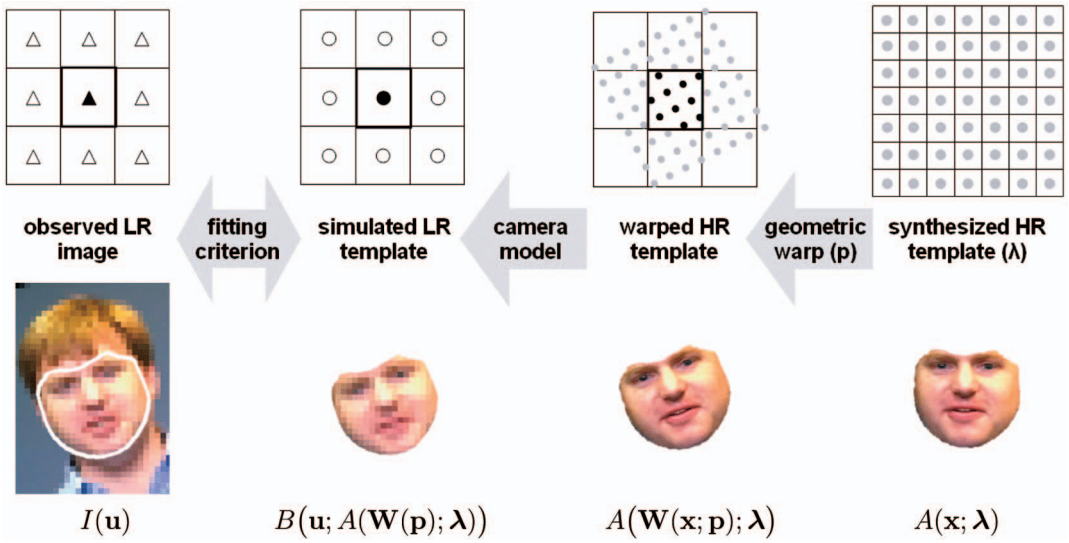


Fig. 5. The Resolution-Aware Fitting (RAF) algorithm simulates the formation of low-resolution images in a digital camera. In contrast to the traditional formulation (Fig. 4), the fitting criterion is defined between observed and simulated image pixels.

as a sum of the mean and linearly weighted basis images, the fitting criterion is

$$\sum_{\mathbf{u}\in \mathrm{dom} I} \left[I(\mathbf{u}) - B\Big(\mathbf{u}; A_0(\mathbf{W}(\mathbf{p})) + \sum_{i=1}^{m} \lambda_i A_i(\mathbf{W}(\mathbf{p}))\Big)\right]^2.$$

Consider the Taylor expansion

$$\sum_{\mathbf{u}\in \mathrm{dom} I} \left[I(\mathbf{u}) - B\Big(\mathbf{u}; A_0(\mathbf{W}(\mathbf{p}+\Delta\mathbf{p})) + \sum_{i=1}^{m} (\lambda_i + \Delta\lambda_i) A_i(\mathbf{W}(\mathbf{p}+\Delta\mathbf{p}))\Big)\right]^2.$$

Ignoring its second-order terms, the fitting criterion is approximately

$$\sum_{\mathbf{u}\in \mathrm{dom} I} \left[I(\mathbf{u}) - B\Big(\mathbf{u}; A_0(\mathbf{W}(\mathbf{p})) + \nabla A_0 \frac{\partial \mathbf{W}}{\partial \mathbf{p}} \Delta\mathbf{p} + \sum_{i=1}^{m} (\lambda_i + \Delta\lambda_i)\Big(A_i(\mathbf{W}(\mathbf{p})) + \nabla A_i \frac{\partial \mathbf{W}}{\partial \mathbf{p}} \Delta\mathbf{p}\Big)\Big)\right]^2.$$

For notational conciseness, denote $n+m$ steepest-descent images as

$$\mathbf{SD}_{sim} = \left[\Big(\nabla A_0 + \sum_{i=1}^{m} \lambda_i \nabla A_i\Big) \frac{\partial \mathbf{W}}{\partial p_1}, \ldots, \Big(\nabla A_0 + \sum_{i=1}^{m} \lambda_i \nabla A_i\Big) \frac{\partial \mathbf{W}}{\partial p_n}, A_1(\mathbf{W}(\mathbf{p})), \ldots, A_m(\mathbf{W}(\mathbf{p}))\right].$$

We can now compactly rewrite the fitting criterion as

$$\sum_{\mathbf{u}\in \mathrm{dom} I} \left[I(\mathbf{u}) - B\left(\mathbf{u}; A_0(\mathbf{W}(\mathbf{p})) + \sum_{i=1}^{m} \lambda_i A_i(\mathbf{W}(\mathbf{p})) - \mathbf{SD}_{sim} \begin{pmatrix} \Delta\mathbf{p} \\ \Delta\boldsymbol{\lambda} \end{pmatrix}\right)\right]^2.$$

Observing that B is a linear operator, the objective function to be minimized is

$$\sum_{\mathbf{u}\in \mathrm{dom} I} \left[I(\mathbf{u}) - B\Big(\mathbf{u}; A_0(\mathbf{W}(\mathbf{p}))\Big) + \sum_{i=1}^{m} \lambda_i B\Big(\mathbf{u}; A_i(\mathbf{W}(\mathbf{p}))\Big) - B(\mathbf{u}; \mathbf{SD}_{sim}) \begin{pmatrix} \Delta\mathbf{p} \\ \Delta\boldsymbol{\lambda} \end{pmatrix}\right]^2,$$

whose minimum is given by

$$\begin{pmatrix} \Delta\mathbf{p} \\ \Delta\boldsymbol{\lambda} \end{pmatrix} = -H_{sim}^{-1} \sum_{\mathbf{u}\in \mathrm{dom} I} B(\mathbf{u}; \mathbf{SD}_{sim}^{\mathrm{T}}) \left[I(\mathbf{u}) - B\Big(\mathbf{u}; A_0(\mathbf{W}(\mathbf{p}))\Big) + \sum_{i=1}^{m} \lambda_i B\Big(\mathbf{u}; A_i(\mathbf{W}(\mathbf{p}))\Big)\right],$$

where H_{sim} is the Hessian with appearance variation:

$$H_{sim} = \sum_{\mathbf{u}\in \mathrm{dom} I} B(\mathbf{u}; \mathbf{SD}_{sim}^{\mathrm{T}}) B(\mathbf{u}; \mathbf{SD}_{sim}).$$

3.4 Quantifying the Improvements of Fit Accuracy by RAF

We will compare the RAF formulation (26) to the traditional formulation in (25). In particular, we will compare the algorithm detailed in Section 3.3.2 (referred to as RAF) with the simultaneous, inverse-compositional algorithm described in [35] (referred to as AAMR-SIM), which optimizes (25). Baker et al. [30], [35] empirically showed that simultaneously solving for the shape and appearance parameters performs better than projecting out the appearance, although at a greater computational cost. Comparing with the simultaneous algorithm in [35] is therefore a fairer comparison than with the project-out algorithm in [32].

We performed two types of experiments: synthetic and real. First, we synthetically downsampled images and compared our fitting results against high-resolution "ground truth" fits. We generated a variety of input test sequences by a range of scaling factors, and measured each algorithm's accuracy at lower input resolutions. These will be presented

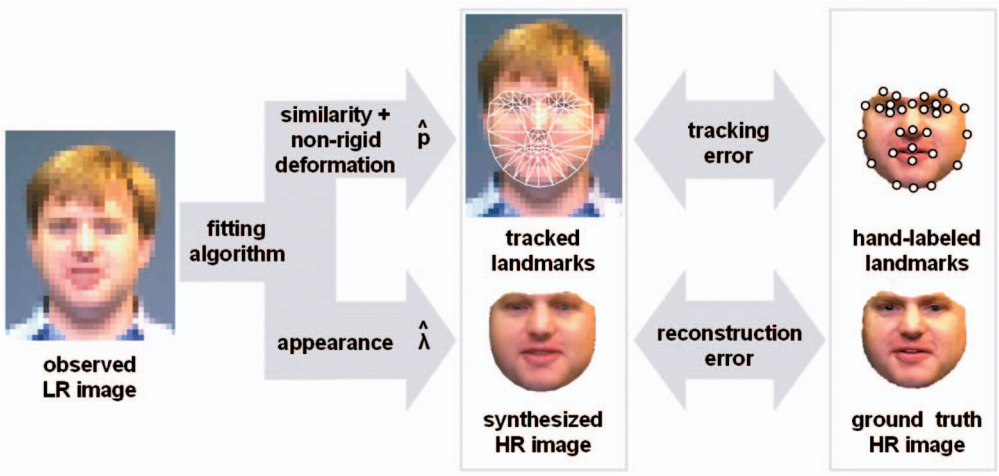


Fig. 6. We define two metrics to compare the fitting accuracy of algorithms. The average landmark tracking error combines the estimation accuracy of the similarity and nonrigid shape parameters. The reconstruction error quantifies how well the underlying high-resolution face could be inferred based only on low-resolution data.

as quantitative results. Second, we ran our algorithms on real low-resolution images to capture their performance under real noise processes. We present these results qualitatively.

Independently of the resolution of a given test sequence, we initialized all algorithms with fitting results at the highest resolution. This allowed us to discard initialization quality as a confounding factor when comparing performances across resolution levels. While manual initialization is reasonable at higher resolutions, it becomes increasingly suboptimal in lower resolutions, jeopardizing the fairness of comparisons across scales. Once in tracking mode, the fitting of each frame was initialized with the parameters of the preceding frame.

3.4.1 Metrics of Fit Accuracy

The most appropriate metric of fit quality depends on applications. For example, in object tracking, only the global pose (i.e., the similarity transform parameters) may be of interest. For lip-reading, nonrigid deformations of a speaker's lips (encoded by a facial AAM's shape coefficients) may carry all the information. If the application requires synthesizing realistic face images, accurate appearance parameter estimates may be of importance.

In the lack of a specific application, we defined two metrics, illustrated in Fig. 6, to summarize the fitting accuracy of the RAF and AAMR-SIM algorithms. The *tracking error* is the average of the positional error of landmarks (such as the corner of nostrils): This error is a combined effect of both similarity transform (scale, rotation, and translation) and nonrigid deformation parameters, as encoded by the estimate $\hat{\mathbf{p}}$. The *reconstruction error*, on the other hand, is computed by comparing the synthesized model instance, parametrized by $\hat{\boldsymbol{\lambda}}$, against the ground truth image in terms of the RMS error of intensities. In addition, we report estimation errors for the coefficients of the top four principal shape and appearance modes.

For all test sequences included in this paper, only the landmark coordinates were available as hand-labeled, high-resolution ground truth data. To infer the ground truth values for the similarity, nonrigid shape, and appearance variables, we ran the AAMR-SIM tracker at the original resolution of the videos and verified its convergence (each landmark's tracking error was smaller than 1 high-resolution pixel). The resulting parameters were then regarded as "ground truth" in subsequent low-resolution tests.

3.4.2 Examples

Before presenting extensive quantitative results, we begin with some examples of our error metrics and their temporal behavior. In reporting Euclidean distance metrics (as in translation parameters or landmark tracking errors), we scale-normalize the estimates so that their numerical values are in high-resolution pixel units. Similarly, we normalize each shape and appearance coefficient according to its mode's variance and report them in units of their standard deviation.

Fig. 7 plots error trajectories of a low-resolution tracking experiment, where the subject's speaking and eye blinking were the major sources of motion. The input sequence was 10 times lower in resolution than the AAM. The error metrics indicate that RAF tracked the face consistently better than AAMR-SIM. To provide further evidence, Fig. 8 shows temporal trajectories of selected variables. Those estimated by AAMR-SIM do not follow the ground truth values and remain mostly constant. In contrast, RAF can track the nonrigid deformations and appearance changes, amounting to a more accurate recovery of the facial expressions. We included this experiment and others in the supplemental video file, which can be found at http://computer.org/tpami/archives.htm and http://www.cs.cmu.edu/~dedeoglu/asymmetry).

3.4.3 Test Set Statistics

It would be impractical to report time trajectories for all our experiments. In the following, we report the temporal mean and standard deviation of the Root Mean Squared (RMS) errors of selected variables. Since lost trackers can easily corrupt these statistics with outliers, we required both trackers to produce valid results (i.e., not have lost track of the face) for a fitting instance to be included in these statistics. This was achieved by visually inspecting all experiments and verifying that faces were tracked reasonably well.

Recall that each tracking experiment was initialized with the highest resolution fitting results. At lower input resolutions, such an optimistic initialization would cause the fitting performance to be overestimated at the beginning. To avoid

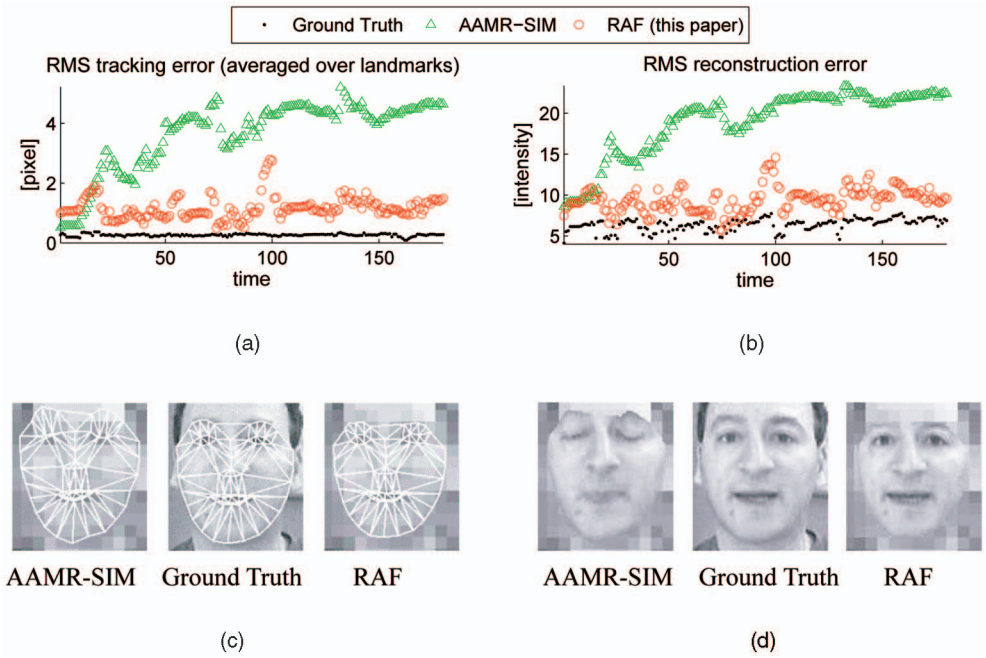


Fig. 7. (a) The landmark tracking and (b) reconstruction error metrics are plotted as a function of time for a 10-fold resolution degraded tracking experiment. Included images (bottom, captured at frame no. 102) display the mesh fits as well as (c) synthesized model images. We overlay the latter onto (d) pixel-replicated low-resolution inputs to demonstrate how well the underlying high-resolution image could be inferred.

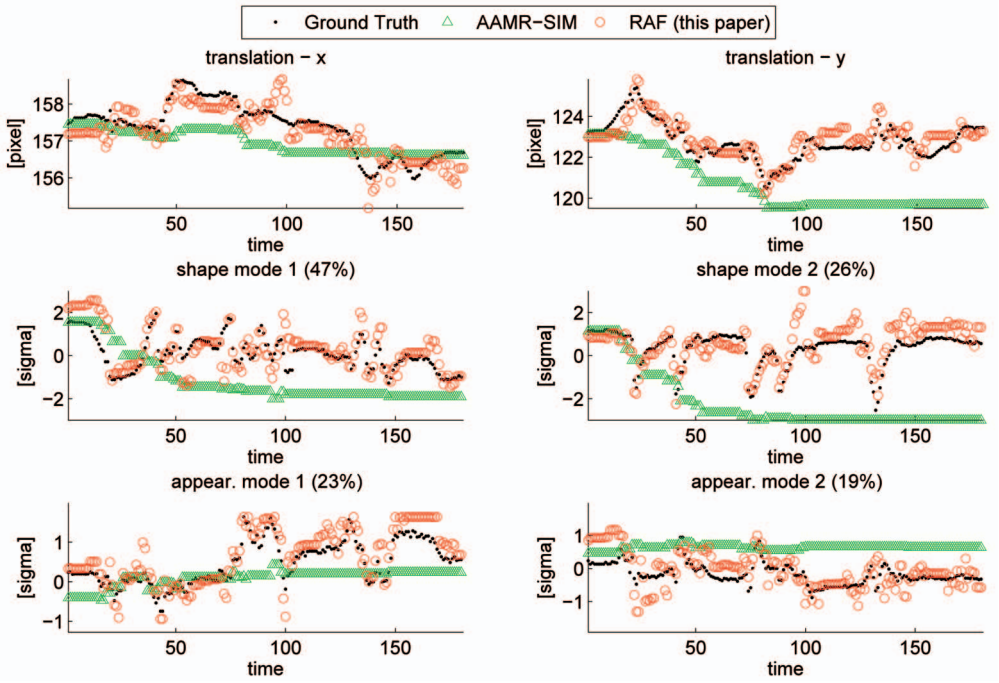


Fig. 8. Selected temporal trajectories are shown for a 10-fold resolution degraded face tracking experiment. As the supplemental video material shows, the main source of motion were the subject's speaking and eye blinking. See Fig. 7 for one example frame of this sequence. The estimates of AAMR-SIM do not follow the ground truth, and remain mostly constant. In contrast, RAF remains close to ground truth in all trajectories, indicating that it is able to extract the underlying facial expressions correctly.

this effect, we discarded the fitting results of the first 20 frames of each sequence.

Fig. 9 compares the AAMR-SIM and RAF algorithms for fitting a single-person AAM. The list at the upper-left corner provides a brief summary of experimental conditions. This AAM was built using 31 training images and was tested on a set of 180. These were 8-bit gray-scale images and the AAM's native resolution was 100×104 pixels. We retained 95 percent of the total variation, yielding 11 shape and 23 appearance principal components.

The plots in Fig. 9 present extensive quantitative comparisons between the fitting algorithms. They are organized to show RMS error metrics as a function of the downscaling factors. Observe how AAMR-SIM and RAF perform equally well at downsampling factor 2. This case corresponds to a minor degradation in resolution, and the fact that both

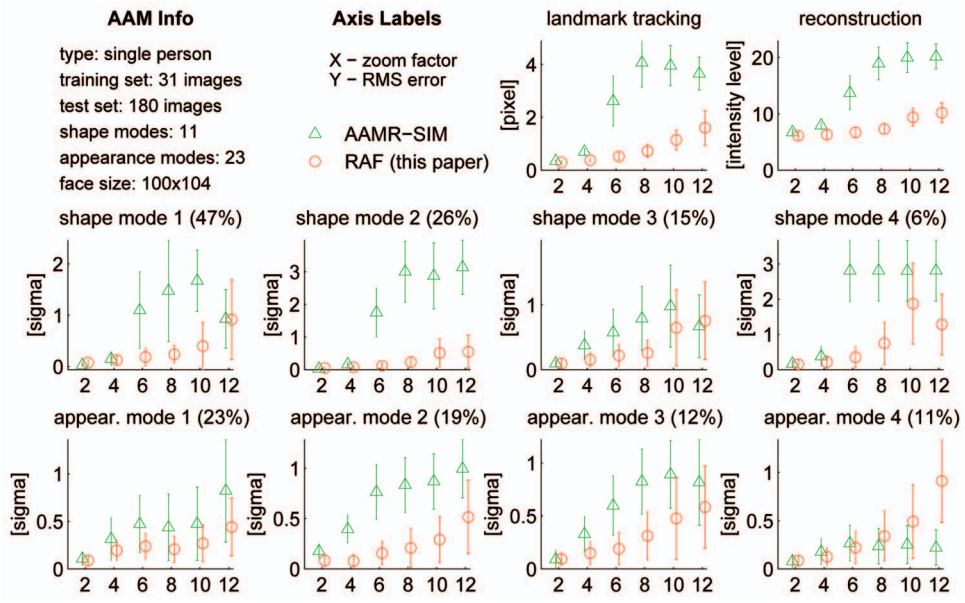


Fig. 9. Quantitative comparison between the AAMR-SIM and RAF algorithms for fitting the single-person AAM to a 180 frame-long sequence. The horizontal axis is the downscaling factor of the input data. Both algorithms perform well at half-resolution, validating the derivation and implementation of RAF. The latter brings substantial improvements across all metrics for downscaling factors 4 and higher. The principal modes are displayed in order of the percent energy (i.e., variation) they capture.

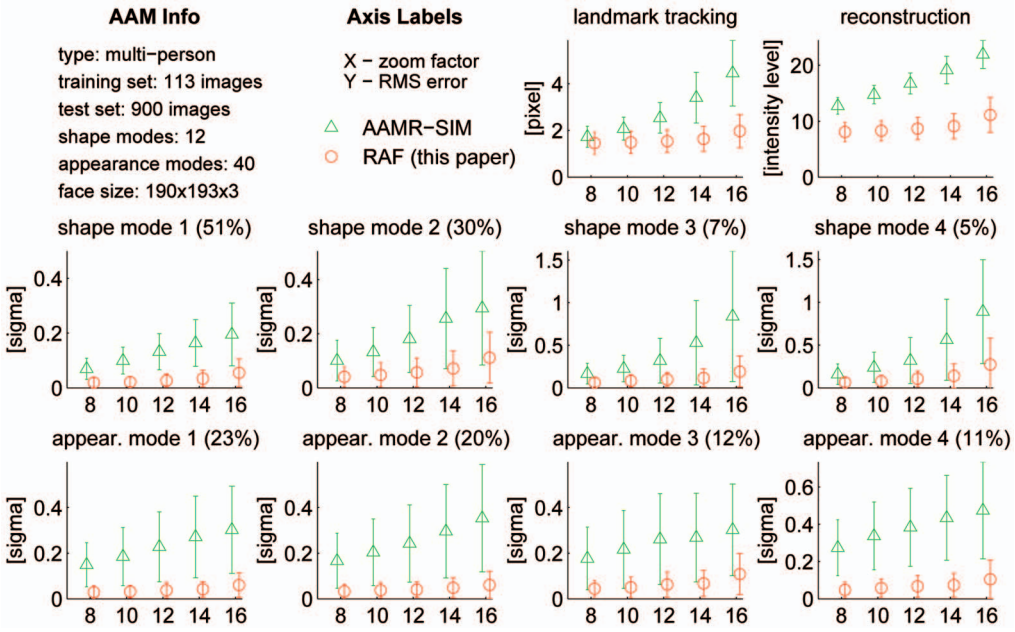


Fig. 10. Quantitative comparison between the AAMR-SIM and RAF algorithms for fitting the multiperson (five subjects) AAM. The horizontal axis is the downscaling factor of the input data. Each reported mean and standard deviation is calculated over 900 frames, comprising 180 frames for each of five subjects. RAF improves the tracking, reconstruction, nonrigid shape, and appearance estimates considerably.

algorithms perform similarly confirms the correctness of our derivations as well as implementations. Starting from downsampling factor 4, RAF brings substantial accuracy improvements over AAMR-SIM across all metrics and variables.

The performance of a model-based method ultimately depends on the quality of the available model. In order to investigate how the AAM fitting accuracy varies with model complexity, we also ran our experiments on a multiperson AAM, which we built using data from five subjects. Details of this AAM are provided in the upper-left corner of Fig. 10, organized in the same fashion as Fig. 9. The multiperson appearance model has almost twice the number of appearance modes compared to the single-person case, indicating a richer subspace being modeled. Again, RAF is consistently superior to AAMR-SIM in accuracy with regard to both tracking and reconstruction.

3.5 Qualitative Results

Once the AAM is fit to a low-resolution face image, we can use its estimated parameters to synthesize a high-resolution reconstruction of the face. This provides a qualitative evaluation of the results. We overlay such reconstructions on pixel-replicated original low-resolution inputs at where the trackers thought the faces were. See the supplemental

Fig. 11. Exemplar subsequence of high-resolution reconstructions, obtained by fitting the single-person AAM. Observe how RAF correctly extracts the eye blink and mouth opening, whereas AAMR-SIM does not. See complete videos at http://www.cs.cmu.edu/~dedeoglu/asymmetry.

video file for tracking sequences, which can be found at http://computer.org/tpami/archives.htm.

Fig. 11 shows every second frame in a sequence of the single-person AAM tracking experiment. Observe that RAF correctly extracts the eye blink and mouth opening, whereas AAMR-SIM does not. Fig. 12 offers a visual alternative for assessing how the trackers degrade with increased downscaling: It displays the single-person AAM results for frame no. 102 across various scales. While RAF can consistently recover the open eyes and mouth, AAMR-SIM's estimates degrade quickly: Starting from downsampling factor 6, the eyes and mouth are first estimated to be half-open, and then totally closed. Similarly, Fig. 13 displays snapshots of different test subjects, all tracked using the multiperson AAM of Fig. 10. In both single and multiperson AAMs, we find the visual reconstruction quality of RAF to be consistently superior to that of AAMR-SIM.

3.6 Qualitative Results on Real Low-Resolution Data

We also compared the two AAM fitting algorithms on real low-resolution videos. Using a Sony DCR-VX2000 camera (15 fps in progressive mode and DV-format compression), we video-taped a particular subject's face at various distances, yielding face heights between 20 and 120 pixels in images. At each camera distance, the subject uttered the sequence "left-right-up-down-smile" and moved her face accordingly. We built a face AAM using 43 high-resolution frames and verified its tracking and reconstruction performance in that resolution. The AAM was $110 \times 114 \times 3$ pixels, with 12 nonrigid shape and 43 appearance modes. We fit this AAM to videos using the AAMR-SIM and RAF algorithms. Initialization was done manually, by scaling and positioning the AAM.

Fig. 14 compares face reconstructions for an eye-blink subsequence. The observed face is 33 pixels high, corresponding to a downscaling factor of about 3. Note the sharpness of RAF reconstructions. In contrast, AAMR-SIM misses the eye-blink and reconstructs blurrier faces.

On all video sequences with downscaling factors 3.5 and higher (where the face height ranged from 30 to 20 pixels), AAMR-SIM consistently lost track of the face. In contrast, RAF kept tracking and reconstructing the face reasonably well. In Fig. 15, we include selected frames of RAF reconstructions. The faces are approximately 22 pixels high. At this resolution, RAF can still recover the underlying facial expression, but the reconstructions start growing unstable.

4 Discussion

4.1 Performance Metrics

Throughout this paper, our discussions have been centered around *accuracy* measures. In certain applications (e.g., nonrigid registration of medical images), criteria such as the repeatability, robustness, and computational efficiency may be as important.

In extremely low resolutions, we found the AAMR-SIM algorithm to be more robust than RAF. Given the smoothing effect of (bilinear) interpolation, this does not seem surprising. While RAF struggles among the many parameter settings which yield almost the same low-resolution images, AAMR-SIM commits to an interpolated high-resolution observation and pursues the fit. In future work, we plan to investigate RAF's robustness properties more closely.

4.2 Implications on the Problem Formulation and Its Optimization

Unless deblurring becomes possible, either the *forward* or the *backward* algorithm will remain biased under relative

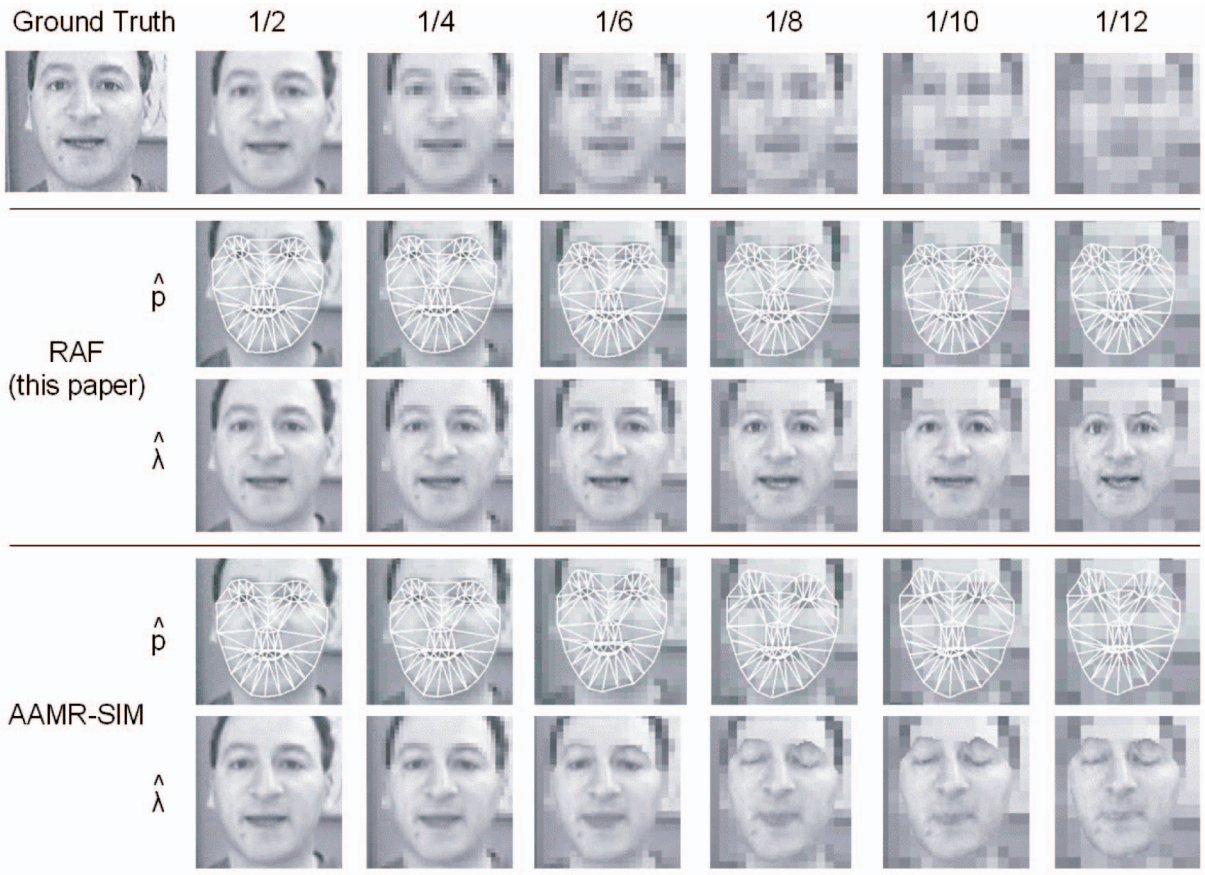


Fig. 12. We compared the AAMR-SIM and RAF algorithms over a range of scales. Lower and lower resolution versions of input frame no. 102 are shown in the top row. While AAMR-SIM degrades quickly, RAF maintains a reasonable estimate of the face.

scaling. By their design, symmetric objective functions of the form (2) can never get rid of this bias. It will always appear in one of the two terms. Nevertheless, if there is not a large scaling, the symmetric formulation is still practical.

From an optimization point of view, constraints on the direction of the warp also have important consequences. Recall how the traditional AAM fitting criterion (25) had conveniently defined the summation over the model template pixels. Since the latter do not change as a function of the input, computational savings become possible: For instance, Matthews and Baker's [32] tracker considers the Taylor expansion for the warp parameters over the model (i.e., AAM appearance basis) and precomputes all associated Jacobians and Hessians. Unfortunately, the RAF formulation of Section 3.3 does not benefit from such precomputations. Implemented in MATLAB, the average fitting time (in tracking mode) of AAMR-SIM is 3 seconds, whereas RAF takes about 10 times longer. One area for future work is to investigate how such savings may be possible.

4.3 Heuristics and Regularization

We only fit nominal-resolution AAMs, independently of how much lower in resolution the observations were. This allowed us to reconstruct faces in high-resolution. A related idea is to construct a scale-space pyramid of AAMs and to model multiple resolutions at once. Due to blur, higher-level (i.e., lower-resolution) AAMs would have more compact appearance models and would therefore be easier to fit. Though this may seem to be an alternative to our approach, comparison *across models* is outside the scope of this paper. In comparing between fitting formulations across a range of resolution degradations, we used exactly the same resolution AAM. Our goal was to solve a given fitting problem *more accurately*, rather than *finding an easier* fitting problem.

We have exclusively dealt with the *formulation* of the image registration problem. The two AAM fitting algorithms compared in Section 3.4 use exactly the same Gauss-Newton minimization method and parameters such as step size, number of iterations, etc. As such, our discussion remains orthogonal to practical search heuristics such as multiresolution, hierarchical, and progressive [4], [6] methods. We can still exploit the advantages of these: For instance, a pyramid style algorithm would increase the robustness of RAF, complementing its accuracy at the bottom level.

Recall that both AAMR-SIM and RAF are Gauss-Newton gradient-descent methods that iteratively update the entire set of AAM parameters. The single and multiperson AAMs have 34 and 52 parameters, respectively, all being estimated simultaneously. Beyond a certain downscaling factor, there are so few face pixels that singularities arise while inverting the Hessians. This practically limited our scaling range to factors of 12 and 16 in the above cases. Parameter scheduling and regularization techniques would help, but such issues are beyond the scope of our paper.

4.4 The Bootstrapping Problem

Cachier and Rey [21] argued that "without any other prior knowledge, the registration problem is symmetric." We claim that the blurry nature of real images breaks this symmetry as

Fig. 13. Selected test frames are shown to visually compare the algorithms for fitting the multiperson AAM. The quantitative improvement in appearance estimates (Fig. 10) has visible effects. Mesh displays are omitted due to a lack of significant difference.

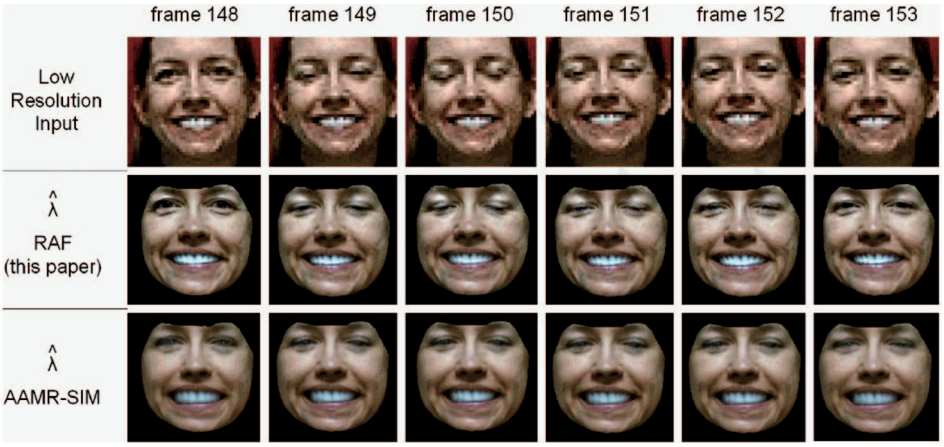


Fig. 14. The top row shows DV-compressed video frames. The face is about 33 pixels high (downscaling factor ~ 3). AAMR-SIM (bottom row) misses the eye-blink and reconstructs overly smooth faces. Indeed, AAMR-SIM fails to track faces any smaller than this size. In contrast, RAF (middle row) infers and reconstructs the underlying facial expression with crisp details.

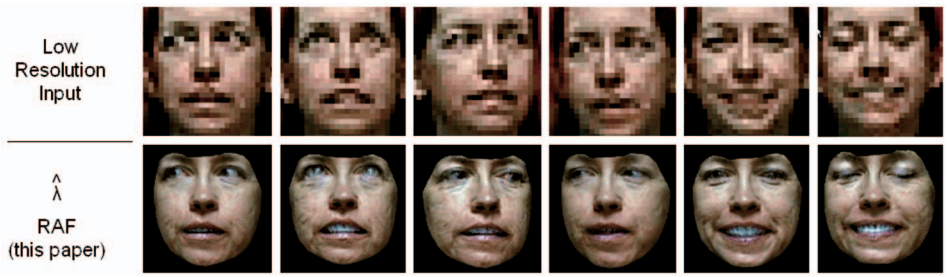


Fig. 15. Selected reconstructions on DV-compressed video. Although the face is only 22 pixels high (downscaling factor ~ 5), RAF can still track the face, recover its expressions, and reconstruct it reasonably well. The temporal jitter and instability observed at this resolution level can be seen in the supplemental video, also available at http://www.cs.cmu.edu/~dedeoglu/asymmetry.

soon as there is relative scaling. Yet, we do not know a priori whether to expect any relative scaling between two images, and if so, which of them ought to be downscaled. This uncertainty raises the question, which warp direction (i.e., *forward* or *backward* algorithm of Section 2) should be employed initially to guess the scaling. The empirical evidence we gathered in the face domain suggests that the bias induced by using the nonoptimal warp algorithm is not big enough to instigate a wrong decision about the direction of the scaling. In other words, we expect both algorithms to be acceptably correct in hinting at the relative scaling, and based on the initial result, we could commit to the *correct* warp direction and obtain unbiased estimates.

4.5 Related Bias Issues in Image Registration Algorithms

Previously, systematic biases of optical flow methods [3], [9], [15], [24], [26] were shown to stem from errors in image gradient estimation or the data-dependence of the noise process [8], [26]. In contrast, we explored a potential bias arising from the resolution-limited nature of real images and showed how this can be taken into account.

In an image matching formulation with point features, Hanson and Morse [16] and Dufournaud et al. [20] observed that their interest points were not invariant to scale. As a remedy, these points were computed for a variety of scale (i.e., blur) levels, which parallels the extra blurring advocated in this paper.

From a practical point of view, we would expect to have difficulties if the two cameras were defocused by different degrees: Since our blur compensation step estimates the amount of necessary blurring from the relative scaling factor, it would not be able to account for the blur accurately. We plan to explore this phenomenon in future work.

5 Conclusion

We argued that the problem of image registration is inherently asymmetric and that ignoring this fact leads to biased estimates. We used a simple yet illuminating scenario starting with an idealized setting wherein the underlying scene radiance field was known. We presented three algorithms (*generative*, *forward*, and *backward*) to compute the ML estimate of the geometric warp which maps between the image coordinate frames. We then investigated which of these algorithms could be used in the absence of scene information. Our analysis exposed the conditions under which *forward* and *backward* algorithms could compute the ML estimate based on the images only, and prescribed a specific blurring step in the presence of relative scaling between images. Such cases turned out to impose a particular warp direction for ensuring unbiased estimates.

Our asymmetry claim is based on the scaling-induced extra blurring that neither a *forward* nor a *backward* algorithm can overcome. It depends upon whichever happens to be warping the lower resolution image onto the higher resolution one. We have shown that an inability to deblur, which is an ill-posed inverse problem, results in a bias that is independent of the assumed observation noise model.

The asymmetry studied in this paper has tangible consequences. We demonstrated and quantified its detrimental effects in human face tracking under low-resolution imaging conditions. The proposed "resolution-aware" formulation indeed resulted in a considerably more accurate tracker.

Finally, our analysis remains applicable to other cases where blur-related discrepancies result not necessarily from camera poses and zoom levels, but from imaging modalities or instrument characteristics.

Appendix

The Equivalence of Forward and Backward Algorithms

Assuming the idealized scenario of Section 2.2, let us express the image warp operations of (11) using point coordinates

$$\hat{\mathbf{W}}_{\mathbf{21}} = \arg\min_{\mathbf{W_{21}}} \int_{\mathbf{y}\in \mathbf{dom} I_1} \Big[S(\hat{\mathbf{W}}_{\mathbf{1S}}(\mathbf{y})) - S\Big(\hat{\mathbf{W}}_{\mathbf{2S}}(\underbrace{\mathbf{W_{12}}(\mathbf{y})}_{\mathbf{z}})\Big)\Big]^2 \mathbf{dy}. \tag{28}$$

We can rewrite the integration in (28) in the domain of I_2 by defining $\mathbf{z} = \mathbf{W_{12}}(\mathbf{y})$. Since $\mathbf{dy} = |J(\mathbf{W_{21}})|\mathbf{dz}$, (28) can be written as

$$\begin{aligned} \hat{\mathbf{W}}_{\mathbf{21}} &= \arg\min_{\mathbf{W_{21}}} \int_{\mathbf{z}\in \mathbf{dom} I_2} \Big[S\Big(\hat{\mathbf{W}}_{\mathbf{1S}}(\mathbf{W_{21}}(\mathbf{z}))\Big) - S\Big(\hat{\mathbf{W}}_{\mathbf{2S}}(\underbrace{\mathbf{W_{12}}(\mathbf{W_{21}}(\mathbf{z}))}_{\mathbf{z}})\Big)\Big]^2 |\, J(\mathbf{W_{21}})|\, \mathbf{dz} \\ &= \arg\min_{\mathbf{W_{21}}} \int_{\mathbf{z}\in \mathbf{dom} I_2} \Big[S\Big(\hat{\mathbf{W}}_{\mathbf{1S}}(\mathbf{W_{21}}(\mathbf{z}))\Big) - S(\hat{\mathbf{W}}_{\mathbf{2S}}(\mathbf{z}))\Big]^2 |\, J(\mathbf{W_{21}})|\, \mathbf{dz}. \end{aligned} \tag{29}$$

Switching back to image warp notation, (29) becomes

$$\hat{\mathbf{W}}_{\mathbf{21}} = \arg\min_{\mathbf{W_{21}}} \int_{\mathbf{z}\in\mathbf{dom}I_2} \Big[warp\big(\underbrace{warp(S;\hat{\mathbf{W}}_{\mathbf{S1}})}_{\hat{I}_1}; \mathbf{W}_{\mathbf{21}}^{-1}\big)(\mathbf{z}) - \underbrace{warp(S;\hat{\mathbf{W}}_{\mathbf{S2}})}_{\hat{I}_2}(\mathbf{z}) \Big]^2 \big| J(\mathbf{W_{21}}) \big| \, \mathbf{dz}. \tag{30}$$

Recalling $\mathbf{W_{12}} = \mathbf{W}_{21}^{-1}$, we observe that the difference between the *forward* (9) and *backward* (30) algorithms' objective functions is the extra Jacobian term $|J(\mathbf{W_{21}})|$ in (30). Since a general homography's Jacobian varies spatially, this term would normally act as a spatial weighting function and influence the minima of the objective function considered. However, in the case of 2D rigid and similarity transformations, the Jacobian would remain constant across the image. Then, the two algorithms would be equivalent and interchangeable.

ACKNOWLEDGMENTS

The authors are grateful to Iain Matthews for kindly providing the AAM Toolbox and sharing his expertise at various stages of implementation and testing. They would also like to thank Jonas August, Yaron Caspi, and members of the CMU misc-reading group for discussions. Finally, they acknowledge the constructive comments of both *IEEE TPAMI* and ECCV 2006 reviewers on earlier versions of this writing [34], [36]. The research described in this paper was supported in part by US Department of Defense contract N41756-03-C4024.

REFERENCES

[1] D.F. Barbe, *Charge-Coupled Devices.* Springer-Verlag, 1980.
[2] B.D. Lucas and T. Kanade, "An Iterative Image Registration Technique with an Application to Stereo Vision," *Proc. Seventh Int'l Joint Conf. Artificial Intelligence,* pp. 674-679, Apr. 1981.
[3] J.K. Kearney, W.B. Thompson, and D.L. Boley, "Optical Flow Estimation: An Error Analysis of Gradient-Based Methods with Local Optimization," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 9, no. 2, pp. 229-244, Mar. 1987.
[4] P. Anandan, "A Computational Framework and an Algorithm for the Measurement of Visual Motion," *Int'l J. Computer Vision,* vol. 2, no. 3, pp. 283-310, Jan. 1989.
[5] C. Casella and R.L. Berger, *Statistical Inference.* Duxbury Press, 1990.
[6] J.R. Bergen, P. Anandan, K.J. Hanna, and R. Hingorani, "Hierarchical Model-Based Motion Estimation," *Proc. European Conf. Computer Vision,* pp. 237-252, May 1992.
[7] L.G. Brown, "A Survey of Image Registration Techniques," *ACM Computing Surveys,* vol. 24, no. 4, pp. 325-376, Dec. 1992.
[8] K. Kanatani, *Statistical Optimization for Geometric Computation: Theory and Practice.* Elsevier, 1996.
[9] J.W. Brandt, "Analysis of Bias in Gradient-Based Optical-Flow Estimation," *Proc. 28th Asilomar Conf. Signals, Systems, and Computers,* vol. 1, pp. 721-725, 1994.
[10] R.C. Hardie, K.J. Barnard, and E.E. Armstrong, "Joint MAP Registration and High Resolution Image Estimation Using a Sequence of Undersampled Images," *IEEE Trans. Image Processing,* vol. 6, no. 12, pp. 1621-1633, Dec. 1997.
[11] M.R. Banham and A.K. Katsaggelos, "Digital Image Restoration," *IEEE Signal Processing Magazine,* vol. 14, no. 2, pp. 24-41, Mar. 1997.
[12] J.B.A. Maintz and M.A. Viergever, "A Survey of Medical Image Registration," *Medical Image Analysis,* vol. 2, no. 1, pp. 1-36, Apr. 1998.
[13] T.F. Cootes, G.J. Edwards, and C.J. Taylor, "Active Appearance Models," *Proc. European Conf. Computer Vision,* vol. 2, pp. 484-498, 1998.
[14] G.J. Edwards, C.J. Taylor, and T.F. Cootes, "Interpreting Face Images Using Active Appearance Models," *Proc. Int'l Conf. Automatic Face and Gesture Recognition,* pp. 300-305, June 1998.
[15] H.-H. Nagel and M. Haag, "Bias-Corrected Optical Flow Estimation for Road Vehicle Tracking," *Proc. Sixth Int'l Conf. Computer Vision,* pp. 1006-1011, Jan. 1998.
[16] B.B. Hansen and B.S. Morse, "Multiscale Image Registration Using Scale Trace Correlation," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* vol. 2, pp. 202-208, 1999.
[17] G.E. Christensen, "Consistent Linear-Elastic Transformations for Image Matching," *Information Processing in Medical Imaging,* pp. 224-237, 1999.
[18] A. Ashburner, J.L.R. Andersson, and K.J. Friston, "High-Dimensional Image Registration Using Symmetric Priors," *NeuroImage,* vol. 9, pp. 619-628, 1999.
[19] M. Irani and P. Anandan, "About Direct Methods," *Proc. Int'l Conf. Vision Algorithms: Theory and Practice,* pp. 267-277, 2000.
[20] Y. Dufournaud, C. Schmid, and R. Horaud, "Matching Images with Different Resolutions," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* pp. 1612-1618, 2000.
[21] P. Cachier and D. Rey, "Symmetrization of the Non-Rigid Registration Problem Using Invertion-Invariant Energies: Application to Multiple Sclerosis," *Proc. Int'l Conf. Medical Image Computing and Computer Assisted Intervention 2000,* pp. 472-481, Oct. 2000.
[22] P.J. Phillips, H. Moon, P.J. Rauss, and S. Rizvi, "The FERET Evaluation Methodology for Face Recognition Algorithms," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 22, no. 10, pp. 1090-1104, Oct. 2000.
[23] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision.* Cambridge Univ. Press, 2000.
[24] C. Fermueller, D. Shulman, and Y. Aloimonos, "The Statistics of Optical Flow," *Computer Vision and Image Understanding,* vol. 82, pp. 1-32, 2001.
[25] T.F. Cootes, G.J. Edwards, and C.J. Taylor, "Active Appearance Models," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 23, no. 6, pp. 681-685, June 2001.
[26] J. Bride and P. Meer, "Registration via Direct Methods: A Statistical Approach," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* vol. 1, pp. 984-989, Dec. 2001.
[27] P. Rogelj and S. Kovacic, "Symmetric Image Registration," *Proc. SPIE,* Feb. 2003.
[28] O. Skrinjar and H. Tagare, "Symmetric, Transitive, Geometric Deformation and Intensity Variation Invariant Nonrigid Image Registration," *Proc. Conf. IEEE Int'l Symp. Biomedical Imaging: Macro to Nano,* vol. 1, pp. 920-923, Apr. 2004.
[29] B. Zitova and J. Flusser, "Image Registration Methods: A Survey," *Image and Vision Computing,* vol. 21, pp. 977-1000, 2003.
[30] S. Baker, R. Gross, and I. Matthews, "Lucas-Kanade 20 Years On: A Unifying Framework: Part 3," Technical Report CMU-RI-TR-03-35, Robotics Inst., Carnegie Mellon Univ., Nov. 2003.
[31] S. Baker and I. Matthews, "Lucas-Kanade 20 Years On: A Unifying Framework," *Int'l J. Computer Vision,* vol. 56, no. 3, pp. 221-255, Mar. 2004.
[32] I. Matthews and S. Baker, "Active Appearance Models Revisited," *Int'l J. Computer Vision,* vol. 60, no. 2, pp. 135-164, Nov. 2004.
[33] J. Modersitzki, *Numerical Methods for Image Registration.* Oxford Univ. Press, 2004.
[34] G. Dedeoğlu and T. Kanade, "On the Source of Asymmetry in Image Registration Problems," Technical Report CMU-RI-TR-05-17, Robotics Inst., Carnegie Mellon Univ., May 2005.
[35] R. Gross, I. Matthews, and S. Baker, "Generic vs. Person Specific Active Appearance Models," *Image and Vision Computing,* vol. 23, no. 11, pp. 1080-1093, Nov. 2005.
[36] G. Dedeoğlu, S. Baker, and T. Kanade, "Resolution-Aware Fitting of Active Appearance Models to Low Resolution Images," *Proc. Ninth European Conf. Computer Vision,* pp. 83-97, 2006.

Göksel Dedeoğlu received the BS degree in control and computer engineering from the Istanbul Technical University in 1997 and the MS degree in computer science from the University of Southern California in 2000. He is a PhD candidate in the Robotics Institute at Carnegie Mellon University. His current research focuses on enhancing low-resolution face videos by means of space-time models and priors. For more details, see his Web page: http://www.cs.cmu.edu/~dedeoglu. He is a student member of the IEEE.

Takeo Kanade received the doctoral degree in electrical engineering from Kyoto University, Japan, in 1974. He is the UA and Helen Whitaker University Professor of Computer Science and Robotics at Carnegie Mellon University. He is also the director of Digital Human Research Center in Tokyo, which he founded in 2001. After holding a faculty position in the Department of Information Science, Kyoto University, he joined Carnegie Mellon University in 1980, where he was the director of the Robotics Institute from 1992 to 2001. Dr. Kanade works in multiple areas of robotics: computer vision, multimedia, manipulators, autonomous mobile robots, medical robotics, and sensors. He has written more than 250 technical papers and reports in these areas and holds more than 20 patents. He has been the principal investigator of more than a dozen major vision and robotics projects at Carnegie Mellon. Dr. Kanade has been elected to the National Academy of Engineering (1997) and the American Academy of Arts and Sciences (2004). He is a fellow of the IEEE, a fellow of the ACM, a founding fellow of the American Association of Artificial Intelligence (AAAI), and the former and founding editor of the *International Journal of Computer Vision*. He has received several awards, including the C&C Award, the Joseph Engelberger Award, the FIT Award, the Allen Newell Research Excellence Award, the JARA Award, the Marr Prize Award, and the Longuet-Higgins Prize. Dr. Kanade has served on advisory or consultant committees for government, industry and university, including the Aeronautics and Space Engineering Board (ASEB) of the National Research Council, NASA's Advanced Technology Advisory Committee, the PITAC Panel for Transforming Healthcare Panel, and the advisory board of the Canadian Institute for Advanced Research.

Simon Baker received the BA degree in Mathematics from Trinity College, Cambridge University, in 1991, the MSc degree in computer science from the University of Edinburgh in 1992, the MA degree in mathematics from Trinity College, Cambridge University, in 1995, and the PhD degree from Columbia University in 1998. He is an associate research professor in the Robotics Institute at Carnegie Mellon University, where he conducts research in computer vision. He is an associate editor of *TPAMI* and will be program chair of CVPR in 2007. His current research interests include face analysis (recognition, tracking, model building, and resolution enhancement), 3D reconstruction, and vision for graphics, vision theory, vision for automotive applications, and projector-camera systems. For more details of his research, see his Web page: http://www.ri.cmu.edu/people/baker_simon.html.