On the Source of Asymmetry in Image Registration Problems

CMU-RI-TR-05-17

Göksel Dedeoğlu and Takeo Kanade

The Robotics Institute
Carnegie Mellon University
Pittsburgh, PA 15232

Abstract. Most image registration problems are formulated in an asymmetric fashion. Given a pair of images, one is implicitly or explicitly regarded as *template*, and warped onto the other to match. 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 for the template, and explain why only the *correct* asymmetric form, with the potential inclusion of a blurring step, can yield unbiased estimators. Our analysis also shows that commonly-used symmetrization/regularization of the warps may actually lead to biased estimators. We experimentally validate our analysis by measuring the bias of L1 and L2 norm motion estimators when the template choice is wrong, and conclude that image registration is inherently an asymmetric problem.

1 Introduction

The problem of image registration underlies a vast array of computer vision applications, such as motion estimation, model-based recognition, and change detection [5, 9, 22]. 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. For instance, the popular "sum of normed differences" objective function [25] takes the form

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

where I_1 and I_2 are images, $\mathbf{W_{12}}$ is the mapping from the coordinate frame of I_1 to that of I_2, and $\mathbf{y}$ is a pixel coordinate in I_1. For $p = 2$ (L2 norm), this amounts to modeling I_1's pixel intensities as i.i.d. Gaussian noise added versions of those of the warped I_2. Consequently, the warp which minimizes (1) is the Maximum-Likelihood (ML) estimate, known to be asymptotically unbiased [4].

Notice that the formulation above is asymmetric: I_2 is regarded as *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. This paper focuses on this seemingly arbitrary choice of the roles, and shows that it leads to biased warp estimates under certain conditions.

The asymmetry issue of (1) has been addressed in prior work [12, 13, 15, 20, 21], where, in an attempt to remove it, the objective functions were explicitly symmetrized[1], yielding

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

In some cases, it was also necessary to enforce the relationship between $\mathbf{W_{12}}$ and $\mathbf{W_{21}}$ through additional *consistency* terms such as

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

which imposed further symmetry constraints on the warps. These past approaches essentially regarded the asymmetry as an opportunity to incorporate more data and regularization priors into the problem at hand.

From our point of view, the scope of relevant work actually goes beyond symmetrization: independent of the *definition* of an objective function, its *numerical optimization* (*i.e.*, the fitting algorithm) may also be treating the two images in an asymmetric fashion. For example, the original Lucas-Kanade algorithm [2] used a Taylor expansion of the warp around its current estimate, yielding

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

[1] Note that simply re-expressing (1) in the domain of I_2 would introduce the Jacobian $|J(\mathbf{W_{21}})|$ as a weighting term. However, the symmetrized form is not necessarily limited to the original noise model. It may instead combine two noise models.

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 [23] proposed an inverse compositional algorithm that performs the expansion on I_1, and minimizes

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

with respect to $\Delta\mathbf{W_{21}}$, resulting in a more efficient fitting. Note that this algorithm warps both images simultaneously, albeit to different degrees.

This paper reveals a fundamental problem overlooked in the past work for symmetrical use of images, and argues that it can only be overcome by using an *asymmetric* form with the *correct* choice of the template, along with a necessary blurring step. We validate our claims experimentally by showing the bias that L1 and L2 norm motion estimators have when the template choice is incorrect. Our results also confirm that the correct template choice and correct warp direction yield unbiased warp estimates.

2 Problem Definition

Let us consider the simplified scenario shown in Fig. 1 (left), 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 [17]. Note that this class of geometric transformation will account for observed images *exactly*. In order to avoid complications arising from non-corresponding image points, we assume that both images have infinite extent and are free of occlusion.

As shown in Fig. 1 (right), 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 homogeneous 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.

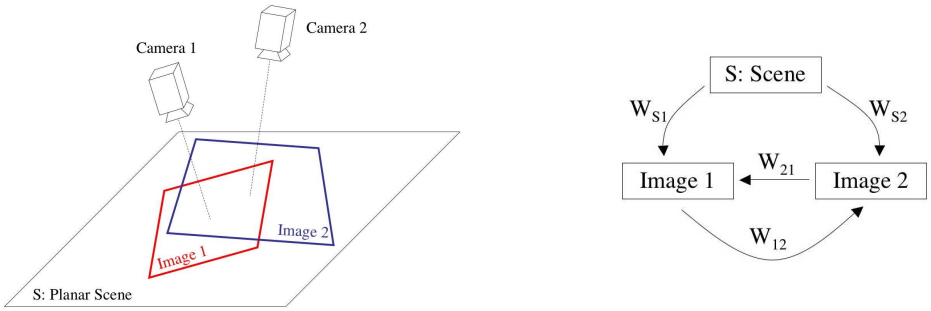

Fig. 1. A planar scene is observed by pinhole cameras (left). Under central projection, scene-to-image and image-to-image transformations (right) are homographies.

We will use two equivalent notations to express the fact that one image is a geometrically transformed version of another. The first one is $I_1(\mathbf{y}) = S\big(\mathbf{W_{1S}}(\mathbf{y})\big)$. 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 $I_1 = \mathbf{W_{S1}}(S)$. This is an abuse of notation for convenience, and refers to an entire domain's transformation. It states that I_1 is the image obtained by transforming every point in the domain of S by $\mathbf{W_{S1}}$. Note the use of $\mathbf{W_{S1}}$ instead of $\mathbf{W_{1S}}$, since the transformed points are in S.

3 Image Registration Scenarios

3.1 Theoretical Case: Ideal Camera and Known Scene

We start our discussion with an idealized scenario. Suppose that we have full knowledge of the underlying scene radiance function S, and both cameras are ideal; their lenses can precisely focus incoming light rays parallel to the optical axis onto the camera's focal point, 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,

$$I_1(\mathbf{y}) = S\big(\mathbf{W_{1S}}(\mathbf{y})\big) + \epsilon(\mathbf{y}) \qquad \forall \mathbf{y} \in I_1, \tag{2}$$

$$I_2(\mathbf{z}) = S\big(\mathbf{W_{2S}}(\mathbf{z})\big) + \epsilon(\mathbf{z}) \qquad \forall \mathbf{z} \in I_2, \tag{3}$$

where $\mathbf{y}$ and $\mathbf{z}$ are points in the domains of I_1 and I_2, respectively. Using the alternative notation, (2) and (3) can be also expressed as

$$I_1(\mathbf{y}) = \mathbf{W_{S1}}(S)(\mathbf{y}) + \epsilon(\mathbf{y}) \qquad \forall \mathbf{y} \in I_1, \tag{4}$$

$$I_2(\mathbf{z}) = \mathbf{W_{S2}}(S)(\mathbf{z}) + \epsilon(\mathbf{z}) \qquad \forall \mathbf{z} \in I_2. \tag{5}$$

In the following, we present three equivalent methods which compute the ML estimate of $\mathbf{W_{12}}$. Given our assumptions at this moment, these algorithms may appear trivial. Nevertheless, the algorithms will be minimally affected while the assumptions will be gradually relaxed throughout the paper. This will allow us to highlight the applicability of these algorithms to different situations.

A1. Compositional Algorithm

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

$$\hat{\mathbf{W}}_{\mathbf{S1}} = \arg\min_{\mathbf{W_{S1}}} \int_{\mathbf{y} \in I_1} \Big[I_1(\mathbf{y}) - \mathbf{W_{S1}}(S)(\mathbf{y}) \Big]^2 \mathbf{dy}. \tag{6}$$

$$\hat{\mathbf{W}}_{\mathbf{S2}} = \arg\min_{\mathbf{W_{S2}}} \int_{\mathbf{z} \in I_2} \Big[I_2(\mathbf{z}) - \mathbf{W_{S2}}(S)(\mathbf{z}) \Big]^2 \mathbf{dz}. \tag{7}$$

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

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

B1. Forward Algorithm

Step 1: Find $\hat{\mathbf{W}}_{\mathbf{S1}}$ and $\hat{\mathbf{W}}_{\mathbf{S2}}$ by (6) and (7).

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 $\mathbf{W}_{\mathbf{12}}$:

$$\hat{\mathbf{W}}_{\mathbf{12}} = \arg\min_{\mathbf{W}_{\mathbf{12}}} \int_{\mathbf{z}\in I_2} \Big[\underbrace{\hat{\mathbf{W}}_{\mathbf{S2}}(S)}_{\hat{I}_2}(\mathbf{z}) - \mathbf{W}_{\mathbf{12}}\big(\underbrace{\hat{\mathbf{W}}_{\mathbf{S1}}(S)}_{\hat{I}_1} \big)(\mathbf{z}) \Big]^2 \mathbf{dz}. \tag{8}$$

By computing $\hat{I}_1 = \hat{\mathbf{W}}_{\mathbf{S1}}(S)$ and $\hat{I}_2 = \hat{\mathbf{W}}_{\mathbf{S2}}(S)$, 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:

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

Note the similarity between (9) and (1): they are both asymmetric, *i.e.*, only one of the images is warped.

C1. Backward Algorithm

Step 1: Find $\hat{\mathbf{W}}_{\mathbf{S1}}$ and $\hat{\mathbf{W}}_{\mathbf{S2}}$ by (6) and (7).

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. This can be done by warping the other ML image:

$$\hat{\mathbf{W}}_{\mathbf{21}} = \arg\min_{\mathbf{W}_{\mathbf{21}}} \int_{\mathbf{y}\in I_1} \Big[\underbrace{\hat{\mathbf{W}}_{\mathbf{S1}}(S)}_{\hat{I}_1}(\mathbf{y}) - \mathbf{W}_{\mathbf{21}}\big(\underbrace{\hat{\mathbf{W}}_{\mathbf{S2}}(S)}_{\hat{I}_2} \big)(\mathbf{y}) \Big]^2 \mathbf{dy}. \tag{10}$$

We 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 Appendix A, we show that the ML warp estimates of all three algorithms would be the same for similarity transforms.

3.2 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 (4) and (5), we have

$$I_1(\mathbf{y}) = B\big(\mathbf{W}_{\mathbf{S1}}(S)\big)(\mathbf{y}) + \epsilon_1(\mathbf{y}) \qquad \forall \mathbf{y} \in I_1, \tag{11}$$

$$I_2(\mathbf{z}) = B\big(\mathbf{W}_{\mathbf{S2}}(S)\big)(\mathbf{z}) + \epsilon_2(\mathbf{z}) \qquad \forall \mathbf{z} \in I_2, \tag{12}$$

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

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

Due to imperfect lenses and density constraints on photo-receptive 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.

In practice, we do not know the scene radiance function S. The estimation of S is an interesting task, and has been posed as a super-resolution problem by [7]. However, we limit the scope of our paper to warp estimation. Let us discuss the three algorithms corresponding to those considered in Section 3.1 for the ideal case.

A2. Compositional Algorithm

Step 1:
$$\hat{\mathbf{W}}_{\mathbf{S1}} = \arg\min_{\mathbf{W_{1S}}} \int\limits_{\mathbf{y}\in I_1} \Big[I_1(\mathbf{y}) - B\big(\mathbf{W_{S1}}(S)\big)(\mathbf{y})\Big]^2 \mathbf{dy}, \tag{13}$$

$$\hat{\mathbf{W}}_{\mathbf{S2}} = \arg\min_{\mathbf{W_{2S}}} \int\limits_{\mathbf{z}\in I_2} \Big[I_2(\mathbf{z}) - B\big(\mathbf{W_{S2}}(S)\big)(\mathbf{z})\Big]^2 \mathbf{dz}. \tag{14}$$

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 S, this algorithm remains impractical.

B2. Forward Algorithm

Step 1: Find $\hat{\mathbf{W}}_{\mathbf{S1}}$ and $\hat{\mathbf{W}}_{\mathbf{S2}}$ by (13) and (14).

Step 2:
$$\hat{\mathbf{W}}_{\mathbf{12}} = \arg\min_{\mathbf{W_{12}}} \int\limits_{\mathbf{z}\in I_2} \Big[\underbrace{B\big(\hat{\mathbf{W}}_{\mathbf{S2}}(S)\big)}_{\hat{I}_2}(\mathbf{z}) - B\Big(\mathbf{W_{12}}\big(\hat{\mathbf{W}}_{\mathbf{S1}}(S)\big)\Big)(\mathbf{z})\Big]^2 \mathbf{dz}. \tag{15}$$

Even though S is unknown, one can still attempt to estimate the objective function (15). Note that the observed image I_2 is the ML estimate for

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

Suppose we denote by T the following "imaging" function

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

Then the image registration problem of (15) becomes

$$\hat{\mathbf{W}}_{\mathbf{12}} = \arg\min_{\mathbf{W_{12}}} \int\limits_{\mathbf{z}\in I_2} \Big[I_2(\mathbf{z}) - T(\mathbf{z})\Big]^2 \mathbf{dz}. \tag{16}$$

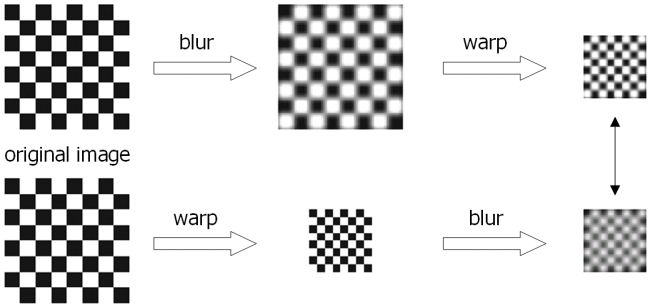

Fig. 2. The order of blurring and geometric warp operations is important: In this example, we used the same Gaussian blur kernel ($\sigma = 2$ pixels) before (top row) or after (bottom row) geometric scaling by a factor of 1/2. Resulting images, shown on the right, differ from each other.

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

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

The observed image I_1 is the ML estimate for $\hat{I}_1 = B\big(\hat{\mathbf{W}}_{\mathbf{S1}}(S)\big)$, and therefore, $T' = \mathbf{W_{12}}\big(I_1\big)$. That is, if we replace T by T' in (16), we would arrive at the commonly used form (1) of objective function in image registration (for $p = 2$):

$$\hat{\mathbf{W}}'_{\mathbf{12}} = \arg\min_{\mathbf{W_{12}}} \int\limits_{\mathbf{z}\in I_2} \Big[I_2(\mathbf{z}) - \mathbf{W_{12}}(I_1)(\mathbf{z}) \Big]^2 \mathbf{dz}. \tag{17}$$

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

While $\hat{\mathbf{W}}'_{\mathbf{12}}$ is biased, there exist, in fact, conditions under which T' can help us compute the unbiased estimate $\hat{\mathbf{W}}_{\mathbf{12}}$. 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 $S' = \hat{\mathbf{W}}_{\mathbf{S1}}(S)$.

$$\begin{aligned} T(\mathbf{x}) &= B\Big(\mathbf{W_{12}}\big(\hat{\mathbf{W}}_{\mathbf{S1}}(S)\big)\Big)(\mathbf{x}) \\ &= B\Big(\mathbf{W_{12}}\big(S'\big)\Big)(\mathbf{x}) \\ &= \int\limits_{\mathbf{w}\in\mathbf{W_{12}}(S')} \mathbf{W_{12}}\big(S'\big)(\mathbf{w}) PSF\big(\mathbf{w}-\mathbf{x}\big)\mathbf{dw}. \end{aligned} \tag{18}$$

On the other hand,

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

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

$$T'(\mathbf{x}) = \int_{\mathbf{w}\in\mathbf{W_{12}}(S')} \mathbf{W_{12}}\big(S'\big)(\mathbf{w}) PSF\big(\underbrace{\mathbf{W_{12}}^{-1}(\mathbf{w}) - \mathbf{W_{12}}^{-1}(\mathbf{x})}_{!}\big)\big|J(\mathbf{W_{12}}^{-1})\big|\mathbf{dw}. \quad (19)$$

Observe that the difference between (18) and (19) is due to the transformation of PSF's argument in (19). Let us consider the case where $\mathbf{W_{12}}$ is 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 (19) is then

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

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, (19) becomes

$$T'(\mathbf{x}) = \int_{\mathbf{w}\in\mathbf{W_{12}}(S')} \mathbf{W_{12}}\big(S'\big)(\mathbf{w}) PSF\big(\frac{\mathbf{w}-\mathbf{x}}{s}\big)\big|J(\mathbf{W_{12}}^{-1})\big|\mathbf{dw}. \quad (20)$$

A comparison of (20) with (18) reveals how the imaging functions T' and T relate to each other. Although they are both obtained by blurring $W_{12}(S')$, the actual blur kernels are different. Imagine that T has the blur kernel $PSF(\cdot)$, shown in the middle of Fig. 3. 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.3, left), 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.3, right), causing T' to be even more blurred than T.

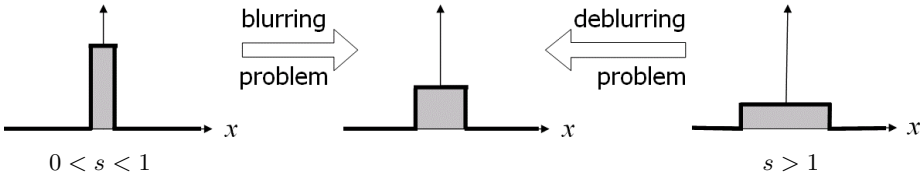

Fig. 3. Given I_1, we can compute $T' = \mathbf{W_{12}}(I_1)$. However, to find the ML warp estimates, we need to evaluate T, which originally results from a convolution operation with the PSF of the camera (middle). Depending on the value of s, estimating T from T' turns out to be a blurring ($0 < s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

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 (17) 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 a typical ill-posed inverse problem, and difficult to solve [8].

Note that the quantities T and T' in (18) and (20) 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.

In cases where W_{12} is not just a similarity transform but a full homography, the inequality between T and T' will still stem from the difference of blur kernels. This time, however, the blur can vary spatially. Under a homography, some parts of the warped image could be subject to compression ($0 < s < 1$) while other parts are being dilated ($s > 1$). This fact calls for a mix of *forward* & *backward* algorithms to be applied to appropriate parts of the image. While such a hybrid algorithm is the theoretically sound way of image registration in general, such perspective distortions do not occur frequently in practice. Thus, we leave the discussion of such a method for future work.

C2. Backward Algorithm

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

4 Experiments

We have found out that the potential asymmetry between *forward* and *backward* registration algorithms is due to the difference in their effective blur kernel widths. Based on our analysis, we expect this difference to become more pronounced as the relative magnification factor between images becomes larger. In order to obtain an unbiased estimate, one must start with the higher-resolution image (I_1 or I_2), blur it appropriately, then warp it onto the lower-resolution one. If the scale-induced blur effect is ignored, or the lower-resolution image is warped onto the higher-resolution one, one should expect the warp estimates to be biased. In this section, we empirically verify these predictions.

4.1 Experimental Procedure

Our analysis shows that when the scene S is not known, any blur in the imaging system will cause the *forward* and *backward* algorithms to differ from each other in the presence of relative scaling. Quantifying the blur effect, however, is not trivial because it is ultimately related to image content: While blurring (*i.e.*, low-pass filtering) visually rich and detailed images would produce a significant effect, 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 it determines the performance of various tracking, recognition, and biometric verification systems. We ran our face-domain experiments on a set of 140 grayscale, frontal face images from the FERET database [16].

In order to quantify the magnitude of image registration bias, we generated synthetic experiments by simulating the image formation process. A real face image, acting as S, was first blurred, then geometrically transformed according to specific warp parameters, and finally resampled to generate images I_1 and I_2. In solving this synthetically generated registration problem, only I_1 and I_2 were used (*i.e.*, unknown scene case).

Since the ground truth warp parameters were known, we could test whether minimizing our objective function gave accurate estimates of the warp. For simplicity, we limited our investigation to similarity transforms with known scaling parameter (s). This left us with three degrees of freedom, namely, translation (t_x, t_y) and rotation (θ), which we considered independently. Having assigned s and θ their ground truth values, we exhaustively searched for the global best values for translation. Similarly, the global minimum for rotation was sought, with s and (t_x, t_y) set to their correct values. These searches were repeated in the neighborhood of their free parameters' ground truth values, and the magnitude of their biases was recorded.

4.2 Testing Conditions

Starting from a `384x256` pixel face image, five different scale reduction parameters (by factors of 1, 2, 4, 8, and 16) were used in generating the test images I_1

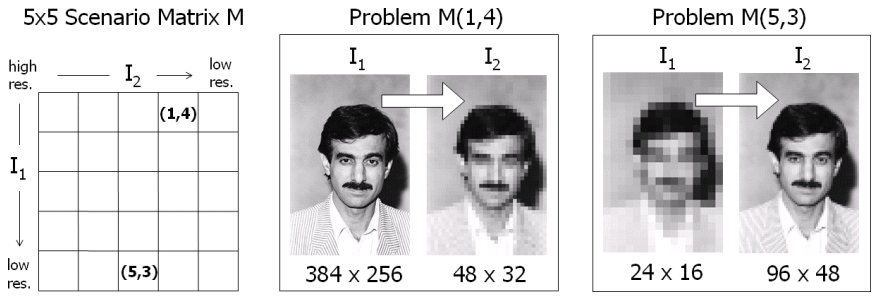

Fig. 4. Our quantitative results are organized in a matrix of registration problem instances (left). Entries of this matrix show the bias statistics of the *forward* algorithm, with blur compensation applied as needed.

and I_2. Without loss of generality, rotation and translation parameters were set to zero. As shown in Fig. 4 (left), this resulted in a `5x5` scenario matrix M of image pairs being registered for each face image. Note that the diagonal elements of this matrix correspond to problem instances where images have the same resolution, whereas off-diagonal elements represent cases where they differ in this respect. Our tests aimed to measure how accurately the ground truth translation and rotation values (*i.e.*, no translation and rotation) could be estimated.

We limited our experiments to the *forward* algorithm, which always warps I_1 onto I_2, regardless of their scale. However, since the full scenario matrix includes all possible pairings, both downscaling and upscaling cases were covered, as exemplified by the instances M(1,4) and M(5,3) in Fig. 4. Bilinear interpolation was used whenever the source image I_1 of the warp was smaller than the destination image I_2, and no deblurring was attempted. In simulating the imaging process, we used a pillbox PSF whose width in scene pixels equaled the integer downscale factor. Similarly, when I_1 was being downscaled, the extra blurring to be applied to T' in (19) was also obtained using a pillbox PSF whose width in I_1 pixels equaled the relative scale factor.

For practical reasons, our "global search" for the best parameter settings was limited to the immediate neighborhood of corresponding ground truth values. For translation, we sampled the interval of $[-0.20, +0.20]$ pixels in 0.01 pixel increments for both t_x and t_y. Similarly, we sampled the $[-2, +2]$ degree interval in 0.1 degree increments for rotation (around the center of the image). Fig. 5 shows example surfaces obtained by sampling translation parameters of L1 and L2 norm objective functions for the problem M(5,3) of Fig. 4, where I_1 is lower in resolution than I_2. As predicted, the global minima of these functions do not lie at the origin, confirming a bias due to the problem formulation (*i.e.*, objective function) itself, rather than the assumed noise or the minimization method.

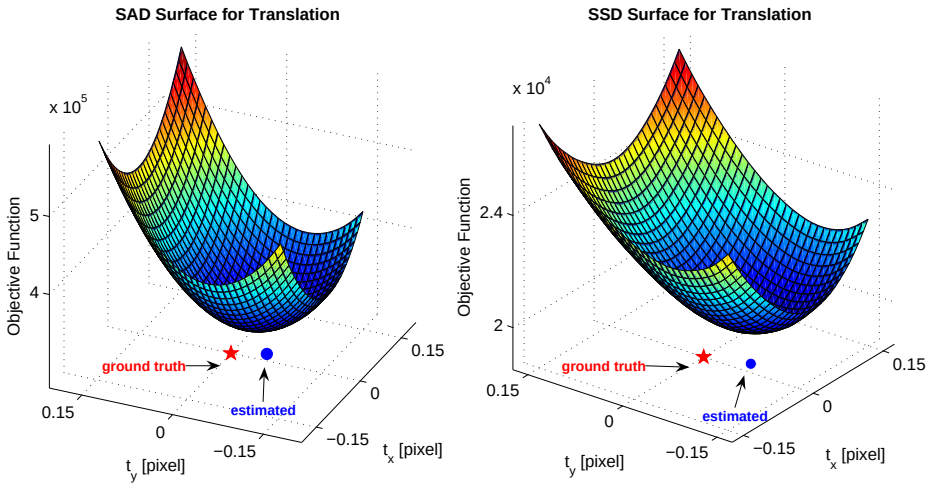

Fig. 5. L1 (left) and L2 (right) norm objective functions for translation parameters of problem instance M(5,3) of Fig. 4. The *forward* algorithm is biased in this case, because its global solution (•) does not coincide with the ground truth (★).

4.3 Results

The *forward* algorithm was investigated for all 25 entries of the scenario matrix, and repeated for 140 face images. Fig. 6 reports means and standard deviations of the computed bias magnitudes, organized in the same matrix form as the scenario matrix. Note that the translation biases are originally measured in I_2 pixel units, because t_x and t_y are added onto I_1 pixel coordinates *after* scaling. Fig. 7 displays selected histograms of translation biases.

The lower triangle of the matrices corresponds to cases where I_1 is lower in resolution than I_2, calling for bilinear interpolation of I_1 during the warp. Confirming our analysis, both translation and rotation parameters are found to be biased. For a given column (*i.e.*, fixed I_2 resolution), we observe that the bias in question gets larger as I_1 is degraded in resolution. This is due to the increased mismatch between T and T' as discussed in Section 3.2, and the fact that the computed objective function increasingly relies on interpolation.

The diagonal and upper triangle of the matrices represent cases where I_1 is equal or higher in resolution than I_2. Following our analysis, we first compensate for the difference between T and T' by blurring I_1 as needed, and then proceed with the geometric warp. As expected, the bias in these cases is empirically found to be zero. For clarity, these entries are not shown.

Translation Bias [pixel]

L1 Norm Obj. Func.

	1	**1/2**	**1/4**	**1/8**	**1/16**
1					
1/2	.01/.01				
1/4	.01/.01	.01/.01			
1/8	.02/.01	.02/.01	.02/.01		
1/16	.03/.02	.04/.02	.04/.02	.04/.02	

L2 Norm Obj. Func.

	1	**1/2**	**1/4**	**1/8**	**1/16**
1					
1/2	.01/.01				
1/4	.01/.01	.02/.01			
1/8	.03/.02	.03/.02	.03/.02		
1/16	.04/.02	.06/.02	.07/.04	.07/.04	

Rotation Bias [degree]

L1 Norm Obj. Func.

	1	**1/2**	**1/4**	**1/8**	**1/16**
1					
1/2	.00/.00				
1/4	.00/.00	.00/.00			
1/8	.00/.02	.00/.00	.00/.00		
1/16	.17/.26	.11/.26	.02/.13	.00/.00	

L2 Norm Obj. Func.

	1	**1/2**	**1/4**	**1/8**	**1/16**
1					
1/2	.00/.00				
1/4	.00/.00	.00/.00			
1/8	.01/.03	.00/.01	.00/.00		
1/16	.14/.19	.11/.21	.04/.14	.01/.07	

Fig. 6. The translation and rotation bias magnitude of the *forward* algorithm organized in the 5x5 scenario matrix form. Rows and columns correspond to scaling factors applied to I_1 and I_2, respectively. Entries are: Mean/Standard Deviation. The diagonal and upper triangle of the matrices are expected, and empirically verified to be zero. For clarity, zero's are not shown in the table. Since the translation parameters are in I_2 pixel units, so are their reported biases. See Appendix B for scale-normalized versions of translation biases.

5 Discussion

5.1 Significance of the Bias in Realistic Cases

Although our discussion has been limited to image-to-image warp estimation, model-based vision algorithms represent a rich domain in which the importance of asymmetry would be pronounced. For instance, [24] presents an efficient tracking algorithm which makes use of Active Appearance Models (AAM) [19] of face shape and appearance. In a surveillance scenario, this algorithm would be fitting high-resolution face models onto low-resolution observations, where relative scales of 1:10 would be common.

In our subsequent work [26], we derived a novel, "resolution-aware" AAM fitting algorithm which took into account the asymmetry due to scaling. We observed important performance differences between the (equivalent of) *forward* and *backward* algorithms, and quantified it for the estimation accuracy of an AAM's shape [2] and appearance parameters. As expected, the algorithm which blurred and warped the high-resolution model onto low-resolution observations was significantly more accurate. This suggests that efficient fitting formulations of [23, 24] may not be readily applicable to low-resolution model-fitting scenarios.

[2] including a few non-rigid modes in addition to the similarity transform

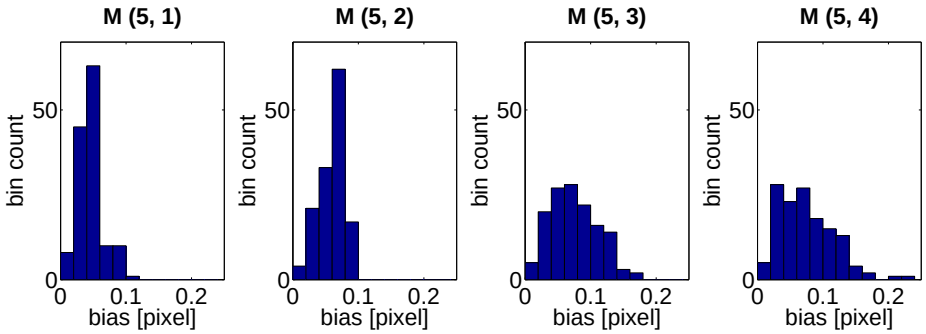

Fig. 7. Translation bias histograms for the L2 Norm objective function, corresponding to the last row (*i.e.*, I_1 scaled by 1/16) of the upper-right scenario matrix M in Fig. 6.

5.2 The Bootstrapping Problem

In [15], authors 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 soon as there is relative scaling (*i.e.*, $|J(W_{12})| \neq 1$). Nonetheless, we agree that *we don't 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 algorithm 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 non-optimal warp algorithm is not big enough to instigate a wrong decision about the direction of the scaling (*i.e.*, whether $0 < s < 1$ or $s > 1$). In other words, we expect both algorithms to be acceptably correct in hinting at the relative scaling, based on which one could commit to the *correct* warp direction and obtain unbiased estimates.

5.3 Related Issues in Image Registration Algorithms

We have discussed the bias problem of registration algorithms. The systematic bias issue in optical flow methods which has been addressed previously [3, 6, 10, 18] stemmed from errors in image gradient estimation. In contrast, we explored a potential bias arising from the *problem formulation* itself. As such, our results remain independent of the limitations of particular numerical implementation and optimization methods.

In an image matching formulation with point features, the authors of [11] and [14] 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 failure mode in future work.

In certain applications (*e.g.*, non-rigid registration of medical images), criteria such as the repeatability and robustness of registration results may be as important as their accuracy. In such cases, trade-offs might be considered.

6 Conclusion

In this paper, we posed the bias problem in image registration in a simple yet illuminating scenario: starting from an idealized setting wherein the underlying scene radiance field S was known, we presented three algorithms (*compositional*, *forward* and *backward*) to compute the ML estimate of the homography which ties the image coordinate frames. We then investigated which of these algorithms could be used in the absence of scene information S. Our analysis exposed the conditions under which *forward* and *backward* algorithms could compute the ML estimate based on I_1 and I_2 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 T' in an attempt to approximate T results in a bias that is independent of the assumed observation noise model: we confirmed this for both L1 and L2 norm objective functions. Furthermore, by showcasing this bias on actual objective function surfaces, we eliminated the possibility of confounding it with potential artifacts of particular optimization methods that would normally search for the best warp parameters.

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.

Acknowledgment
Authors are grateful to Jonas August, Simon Baker, Gerard Barnes, and the CMU misc-reading group for helpful discussions and suggestions.

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. of the 7th Int. Joint Conf. on Artificial Intelligence*, April, 1981, pp. 674-679.
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. on Pattern Analysis and Machine Intelligence*, Vol. 9, No. 2, March, 1987, pp. 229-244.
4. C. Casella, R.L. Berger: Statistical Inference. Duxbury Press, Belmont, CA, 1990.

5. L.G. Brown: A Survey of Image Registration Techniques. *ACM Computing Surveys*, Vol. 24, No. 4, December, 1992, pp. 325-376.
6. J.W. Brandt: Analysis of bias in gradient-based optical-flow estimation. *Proc. 28th Annual Asilomar Conf. on Signals, Systems, and Computers*, Vol. 1, pp. 721-725.
7. 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. on Image Processing*, Vol. 6, No. 12, Dec. 1997, pp. 1621-1633.
8. M.R. Banham and A.K. Katsaggelos: Digital Image Restoration. *IEEE Signal Processing Magazine*, Vol. 14, no. 2, Mar. 1997, pp. 24-41.
9. J.B.A. Maintz and M.A. Viergever: A survey of medical image registration. *Medical Image Analysis*, Vol. 2, No. 1, April, 1998, pp. 1-36.
10. H.-H. Nagel and M. Haag: Bias-corrected optical flow estimation for road vehicle tracking. *Proc. of Sixth Inter. Conf. on Computer Vision*, Jan. 1998, pp. 1006-1011.
11. B.B. Hansen and B.S. Morse: Multiscale Image Registration Using Scale Trace Correlation. *IEEE Conf. on Comp. Vision and Pattern Recognition*, Vol. 2, June, 1999, pp.202-208.
12. G.E. Christensen: Consistent Linear-Elastic Transformations for Image Matching. *Information Proces. in Medical Imaging*, LCNS 1613, Springer, 1999, pp. 224-237.
13. A. Ashburner, J.L.R. Andersson, and K.J. Friston: High-Dimensional Image Registration Using Symmetric Priors. *NeuroImage*, Vol. 9, 1999, pp. 619-628.
14. Y. Dufournaud, C. Schmid, and R. Horaud: Matching Images with Different Resolutions. *IEEE Conf. on Comp. Vision and Pattern Recognition*, 2000, pp. 1612-1618.
15. P. Cachier and D. Rey: Symmetrization of the Non-Rigid Registration Problem using Invertion-Invariant Energies: Application to Multiple Sclerosis. *Proc. of MICCAI 2000*, October, pp. 472-481.
16. P.J. Phillips, H. Moon, P.J. Rauss, and S. Rizvi: The FERET Evaluation Methodology for Face Recognition Algorithms. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 22, No. 10, October 2000.
17. R. Hartley and A. Zisserman: Multiple View Geometry in Computer Vision. Cambridge University Press, 2000.
18. C. Fermueller, D. Shulman and Y. Aloimonos: The Statistics of Optical Flow. *Computer Vision and Image Understanding*, Vol. 82, 2001, pp. 1-32.
19. T.F. Cootes, G.J. Edwards, and C.J. Taylor: Active Appearance Models. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 23 (6), 2001, pp. 681-685.
20. P. Rogelj and S. Kovacic: Symmetric Image Registration. *Proc. of SPIE*, Vol. 5032, Medical Imaging 2003: Image Processing, February 2003.
21. O. Skrinjar and H. Tagare: Symmetric, transitive, geometric deformation and intensity variation invariant nonrigid image registration. *IEEE Int. Symp. on Biomedical Imaging: Macro to Nano, 2004*. Vol. 1, April, 2004, pp. 920-923.
22. B. Zitova and J. Flusser: Image Registration Methods: A Survey. *Image and Vision Computing* Vol. 21, 2003, pp. 977-1000.
23. S. Baker and I. Matthews: Lucas-Kanade 20 Years On: A Unifying Framework. *Int. Journal of Computer Vision*, Vol. 56, No. 3, March, 2004, pp. 221-255.
24. I. Matthews and S. Baker: Active Appearance Models Revisited. *Int. Journal of Computer Vision*, Vol. 60, No. 2, November, 2004, pp. 135-164.
25. J. Modersitzki: Numerical Methods for Image Registration. Oxford U. Press, 2004.
26. G. Dedeoglu and S. Baker and T. Kanade: Resolution-Aware Fitting of Active Appearance Models to Low Resolution Images. *Proc. of European Conf. on Computer Vision*, May, 2006, to appear.

A The Equivalence of Forward and Backward Algorithms

Assuming the idealized scenario of Section 3.1, let us express the image warp operations of (10) using point coordinates

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

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

$$\begin{aligned}\hat{\mathbf{W}}_{\mathbf{21}} &= \arg\min_{\mathbf{W_{21}}} \int_{\mathbf{z}\in I_2} \Big[S\Big(\hat{\mathbf{W}}_{\mathbf{1S}}(\mathbf{W_{21}}(\mathbf{z}))\Big) - S\Big(\hat{\mathbf{W}}_{\mathbf{2S}}(\underbrace{\mathbf{W_{12}}(\mathbf{W_{21}}(\mathbf{z}))}_{\mathbf{z}})\Big)\Big]^2 |J(\mathbf{W_{21}})|\mathbf{dz} \\ &= \arg\min_{\mathbf{W_{21}}} \int_{\mathbf{z}\in I_2} \Big[S\Big(\hat{\mathbf{W}}_{\mathbf{1S}}(\mathbf{W_{21}}(\mathbf{z}))\Big) - S\big(\hat{\mathbf{W}}_{\mathbf{2S}}(\mathbf{z})\big)\Big]^2 |J(\mathbf{W_{21}})|\mathbf{dz}. \end{aligned} \tag{22}$$

Switching back to image warp notation, (22) becomes

$$\hat{\mathbf{W}}_{\mathbf{21}} = \arg\min_{\mathbf{W_{21}}} \int_{\mathbf{z}\in I_2} \Big[\mathbf{W}_{21}^{-1}(\underbrace{\hat{\mathbf{W}}_{\mathbf{S1}}(S)}_{\hat{I}_1})(\mathbf{z}) - \underbrace{\hat{\mathbf{W}}_{\mathbf{S2}}(S)}_{\hat{I}_2}(\mathbf{z})\Big]^2 |J(\mathbf{W_{21}})|\mathbf{dz}. \tag{23}$$

Recalling $\mathbf{W_{12}} = \mathbf{W_{21}^{-1}}$, we observe that the difference between the *forward* (8) and *backward* (23) algorithms' objective functions is the extra Jacobian term $|J(\mathbf{W_{21}})|$ in (23). 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 practically common cases of rigid and similarity transformations, the Jacobian would remain constant across the image. Then, the two algorithms would be equivalent, and interchangeable.

B Scale-Normalized Translation Biases

As indicated in Section 4.3, the translation bias results of Fig. 6 are reported in I_2's pixel units. However, since I_2 has a different resolution in every column, the entries of the matrix are not directly comparable. In Fig. 8, we replicate these bias values in a common (highest-resolution) scale.

Translation Bias [pixel]

L1 Norm Obj. Func.

	1	**1/2**	**1/4**	**1/8**	**1/16**
1					
1/2	.01/.01				
1/4	.01/.01	.03/.02			
1/8	.02/.01	.04/.03	.09/.05		
1/16	.03/.02	.08/.04	.17/.09	.36/.19	

L2 Norm Obj. Func.

	1	**1/2**	**1/4**	**1/8**	**1/16**
1					
1/2	.01/.01				
1/4	.01/.01	.03/.02			
1/8	.03/.03	.06/.04	.13/.07		
1/16	.04/.02	.12/.04	.30/.15	.59/.33	

Fig. 8. Scale-normalized translation bias values (*c.f.* Fig. 6)