PARAMETERISATION AND PROBABILITY IN IMAGE ALIGNMENT

Nicholas Molton, Andrew Davison, Ian Reid

Department of Engineering Science
University of Oxford
Parks Road, Oxford OX1 3PJ,
UK

ABSTRACT

In this paper we investigate extending the gradient-based *inverse compositional* image alignment method described by Baker and Matthews [1] by formulating the alignment process probabilistically using Bayes rule to obtain a posterior estimate of the alignment. The probabilistic formulation makes use of prior statistical information on the aligning function parameters, and we investigate the use of arbitrary parameterisation of this function to match the physical system generating the warp.

We compare the probabilistic method to the standard inverse compositional algorithm by using affine alignment to track locally planar image regions through image sequences in real-time. We show that the stabilising effect of probabilistic alignment gives more reliable results with little effect on alignment speed. For this application we also find that the choice of warp parameterisation is significant in its own right, as we get much better results from the standard inverse compositional algorithm with a more physically motivated affine parameterisation.

1. BACKGROUND TO IMAGE ALIGNMENT

Image alignment as described here is the process of bringing into alignment two image regions by finding the parameters of a known function relating position in one region to position in the other, where the parameters are initially only approximately known.

Alignment is useful for any application where pixels in a local area are expected to move in way which can be modelled by a function of position; for example, for tracking rigid objects of known surface shape moving in 3D, tracking the images from a rotating camera, or tracking non-rigid objects which deform in a way which can realistically be modelled.

The image alignment problem has been studied for a number of years. Early work using a gradient method was done by Lucas and Kanade [2], and many authors have since looked at modifications to this [3][4][5][1][6]. There are also a range of methods which perform image alignment in very different ways for different applications, for example, optic flow [7], look up techniques [8], or feature matching [9]

1.1. Inverse compositional image alignment

The work described here is based on the inverse compositional image alignment algorithm of Baker and Matthews [1], which is a more efficient formulation of the Lucas and Kanade method. Using Baker and Matthews' notation, the method aims to align an image $\mathtt{I}$ with a template image $\mathtt{T}$, as illustrated in Figure 1. $I(\mathbf{x})$, and $T(\mathbf{x})$ represent the intensity in images $\mathtt{I}$ and $\mathtt{T}$ at image position $\mathbf{x}$. The alignment function $\mathtt{W}(\mathbf{x};\mathbf{p})$ maps an image position $\mathbf{x}$ to a new position $\mathbf{x}'$, and is a function of the image position $\mathbf{x}$ and a set of warp parameters $\mathbf{p}$. The method aims to minimise the following cost function with respect to $\Delta\mathbf{p}$:

$$\sum_i \left[T(\mathtt{W}(\mathbf{x};\Delta\mathbf{p})) - I(\mathtt{W}(\mathbf{x};\mathbf{p}))\right]^2, \qquad (1)$$

where $\mathbf{p}$ corresponds to the current estimate of the set of warp parameters needed to bring the two images into alignment, and variation of $\Delta\mathbf{p}$ is used to obtain the most optimal alignment from the image data. The summation is done on a set of pixels in the region of interest. This might be all the pixels in an area, or the subset of these for which the image intensity gradient is significant. The warp estimate is then updated as follows:

$$\mathtt{W}(\mathbf{x};\mathbf{p})_{new} = \mathtt{W}(\mathbf{x};\mathbf{p})_{previous} \circ \mathtt{W}(\mathbf{x};\Delta\mathbf{p})^{-1}; \qquad (2)$$

that is the new warp is the total effect of applying the old warp, followed by the inverse of the warp defined by $\Delta\mathbf{p}$. The minimisation is done by performing a first order Taylor expansion on Equation 1 at the point $\mathbf{x}$ to give:

$$\sum_i \left[T(\mathtt{W}(\mathbf{x};\vec{0})) + \nabla T \frac{\partial \mathtt{W}}{\partial \mathbf{p}} \Delta\mathbf{p} - I(\mathtt{W}(\mathbf{x};\mathbf{p}))\right]^2, \qquad (3)$$

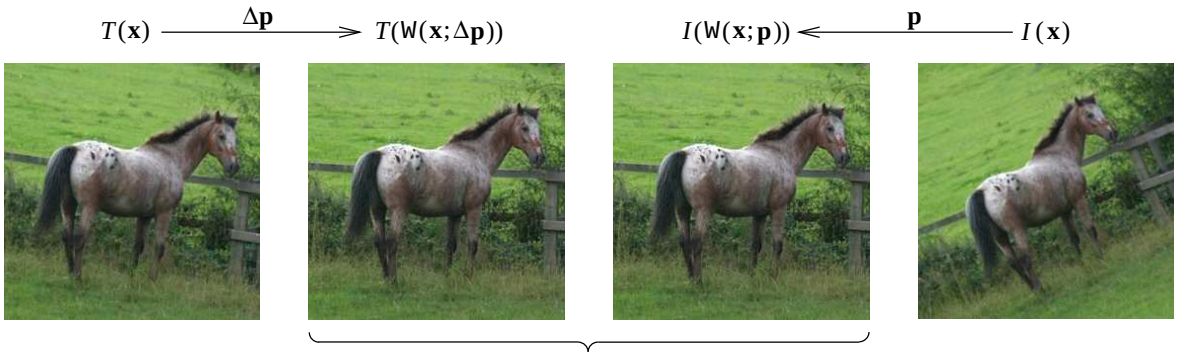

Fig. 1. Inverse Compositional image alignment. The warp $\mathtt{W}(\mathbf{x};\mathbf{p})$ is the current estimate of the warp required for alignment. The warp $\mathtt{W}(\mathbf{x};\Delta\mathbf{p})$ is a small adjustment warp calculated in the alignment algorithm. Because this adjustment is calculated from gradient properties of T, which are constant for all images I, the algorithm is computationally efficient.

where the Jacobian $\frac{\partial \mathtt{W}}{\partial \mathbf{p}}$ is evaluated at $\mathbf{p} = \vec{0}$ and the current value of $\mathbf{x}$. It is assumed that the parameter vector $\vec{0}$ maps points to themselves. Differentiating and setting the result to zero gives (see [1] for more details):

$$\Delta\mathbf{p} = \mathtt{H}^{-1} \sum_i \left[\nabla T \frac{\partial \mathtt{W}}{\partial \mathbf{p}}\right]^T [I(\mathtt{W}(\mathbf{x};\mathbf{p})) - T(\mathbf{x})], \quad (4)$$

where

$$\mathtt{H} = \sum_i \left[\nabla T \frac{\partial \mathtt{W}}{\partial \mathbf{p}}\right]^T \left[\nabla T \frac{\partial \mathtt{W}}{\partial \mathbf{p}}\right].$$

1.2. The need for probabilistic image alignment

The inverse compositional algorithm, and other-non probabilistic image alignment algorithms often become unstable when the number of parameters in the warp rises above about three or four. A number of authors note instability when using an affine warp [3][5]. We hope to improve on the stability of the method by solving the alignment problem probabilistically, taking into account uncertainty in the image difference measurements and prior information about the warp parameters.

The standard inverse compositional algorithm uses a prior estimate of the warp (defined by $\mathbf{p}$) as the starting point for alignment. We aim to also make use of a measure of the uncertainty in this prior estimate. This tells us which components should change in preference to others during alignment and gives us a measure of the overall change which is reasonable in any alignment step.

Since prior information will often come from a model of a physical system, with its own set of parameters, we wish to use the same parameters to describe the warp used in alignment. This will ensure that the warping function is correct, and that uncertainty information can be passed directly between the alignment process and the physical model. In general this will result in an arbitrarily complex function to describe the warp, and the consequence of this is discussed in Section 2.

We formulate image alignment as the calculation of the posterior distribution of $\Delta\mathbf{p}$ calculated through Bayes rule. This allows prior information and sources of uncertainty in the inverse compositional measurement $(I(\mathtt{W}(\mathbf{x};\mathbf{p})) - T(\mathbf{x}))$ to be taken into account. As well as estimating the warp, this method also gives a confidence measure in it, which makes the result more useful to other processes.

2. PARAMETERISATION OF THE IMAGE WARP

The transform relating views of an object can often be described by a simple and compact equation. For example, a 2D homography describes the transform relating views of a planar surface moving in 3D, or views from a rotating camera, or many common non-rigid 2D transformations such as shearing, stretching, or scaling. For a particular application we assume that a possibly over-general, but algebraically simple expression like this is known. This expression, $\mathtt{W}_s$, is a function of image position $\mathbf{x}$, and a set of parameters $\mathbf{p}_s$:

$$\mathbf{x}' = \mathtt{W}_s(\mathbf{x}, \mathbf{p}_s).$$

Now suppose we would actually prefer to describe the transformation using a different parameterisation, W, which better reflects the physical system being studied:

$$\mathbf{x}' = \mathtt{W}(\mathbf{x}, \mathbf{p}).$$

For example we might be tracking a planar surface on a mechanism moving in a restricted way in 3D, or tracking images from a rotating camera. The transformation in these cases is a homography, but a more specialised parameterisation whose parameters also have physical meaning is possible.

In general the warp $\mathtt{W}$ can be quite complex, which could make the inverse compositional image alignment method computationally expensive if it is used directly. If, however, we can find the simplest algebraic expression which is general enough to describe the system ($\mathtt{W}_s$), and write its parameters as a function of the parameters of $\mathtt{W}$, the effect on the inverse compositional algorithm should be minimal. If the function relating the two parameter vectors is:

$$\mathbf{p}_s = f(\mathbf{p}),$$

where $\mathbf{p}$ is the parameter vector of $\mathtt{W}$ and $\mathbf{p}_s$ is the parameter vector of $\mathtt{W}_s$. The warp function $\mathtt{W}$ then becomes:

$$\mathbf{x}' = \mathtt{W}(\mathbf{x}, \mathbf{p}) = \mathtt{W}_s(\mathbf{x}, \mathbf{p}_s) = \mathtt{W}_s(\mathbf{x}, f(\mathbf{p})),$$

and the Jacobian of the warp transform becomes:

$$\frac{\partial \mathtt{W}}{\partial \mathbf{p}}(\mathbf{x}, \mathbf{p}) = \frac{\partial \mathtt{W}_s}{\partial \mathbf{p}_s}(\mathbf{x}, \mathbf{p}_s)\frac{\partial \mathbf{p}_s}{\partial \mathbf{p}}(\mathbf{p}) = \frac{\partial \mathtt{W}_s}{\partial \mathbf{p}_s}(\mathbf{x}, f(\mathbf{p}))\frac{\partial f(\mathbf{p})}{\partial \mathbf{p}}(\mathbf{p})$$

The inverse compositional algorithm needs to evaluate these values for every pixel being used to calculate the alignment but the calculation has been broken down into the evaluation of $f(\mathbf{p})$ and $\frac{\partial f(\mathbf{p})}{\partial \mathbf{p}}(\mathbf{p})$, which are potentially slow but need only be done once per iteration, and the evaluation of $\mathtt{W}_s(\mathbf{x}, f(\mathbf{p}))$ and $\frac{\partial \mathtt{W}_s}{\partial \mathbf{p}_s}(\mathbf{x}, f(\mathbf{p}))$, which are relatively simple operations, at every pixel. Therefore the warping function can be reparameterised with negligible cost to the inverse compositional algorithm.

One possible complication is that the inverse compositional image alignment algorithm linearises the warp function around the zero warp position, so parameterisations which are excessively nonlinear in this region might be expected to work less well.

We tested the usefulness of parameterisation on its own by using an inverse compositional tracker to align the images coming from a rotating camera, and comparing the result to measurements coming from a gyroscopic sensor. For this application the parameterisation constrains the general 2D homography ($\mathtt{W}_s$), with eight degrees of freedom, to the conjugate rotation corresponding to our calibrated, fixed focal length camera, which has only three degrees of freedom. The resulting rotation tracker was able to very reliably track the motion of a rotating camera in real-time [10].

3. PROBABILISTIC INVERSE COMPOSITIONAL IMAGE ALIGNMENT

The inverse compositional algorithm calculates a least squares estimate of a state $\Delta\mathbf{p}$ from a series of measurements (one for each pixel) which are a function of the state:

$$z = I(\mathtt{W}(\mathbf{x}; \mathbf{p})) - T(\mathbf{x}) \approx \nabla T \frac{\partial \mathtt{W}}{\partial \mathbf{p}} \Delta\mathbf{p} \qquad (5)$$

We wish to calculate a better estimate of $\Delta\mathbf{p}$ by specifying uncertainty in the measurement equation, and prior information on $\Delta\mathbf{p}$, and including them in the calculation. This estimate is the posterior distribution of $\Delta\mathbf{p}$, given by Bayes rule:

$$p(\Delta\mathbf{p}|z) = \frac{p(z|\Delta\mathbf{p})p(\Delta\mathbf{p})}{p(z)} = \frac{p(z|\Delta\mathbf{p})p(\Delta\mathbf{p})}{\int_{-\infty}^{\infty} p(z|\Delta\mathbf{p})p(\Delta\mathbf{p})d\Delta\mathbf{p}} \qquad (6)$$

3.1. A likelihood function for $\Delta\mathbf{p}$

A more accurate version of Equation 5 includes terms representing uncertainty in the measurement:

$$z = \nabla T \frac{\partial \mathtt{W}}{\partial \mathbf{p}} \Delta\mathbf{p} + n_1 + \nabla \mathtt{T}\, \mathbf{n}_2 + \nabla \mathtt{T}\, \frac{\partial \mathtt{W}}{\partial \mathbf{p}} \mathbf{n}_3, \qquad (7)$$

where the terms n_1, $\mathbf{n}_2$, and $\mathbf{n}_3$ are zero mean Gaussian random variables of dimensionality 1, 2, and n (the size of the warp parameter vector) respectively. Between them they are intended to represent uncertainties which are identical at each pixel (for example, camera noise), uncertainty which increases when contrast is high (for example, aliasing effects), and uncertainties which increase with distance from the patch centre (for example errors introduced through linearisation in the measurement equation). This choice of uncertainty model is discussed in more detail in [10]. For simplicity we avoid making the uncertainty terms (the n terms) a function of either $\mathbf{p}$ or $\Delta\mathbf{p}$. The likelihood function $p(z|\Delta\mathbf{p})$ following from this uncertainty model is:

$$p(z|\Delta\mathbf{p}) \quad = \quad \frac{1}{\sqrt{2\pi\sigma_z^2}} e^{-\frac{1}{2}\sigma_z^{-2}(z - \frac{\partial \mathtt{T}}{\partial \mathbf{p}}\Delta\mathbf{p})^2} \qquad (8)$$

where,

$$\sigma_z^2 = \sigma_1^2 + \nabla T \Lambda_2 \nabla T^T + \nabla T \frac{\partial \mathtt{W}}{\partial \mathbf{p}} \Lambda_3 \frac{\partial \mathtt{W}}{\partial \mathbf{p}}^T \nabla T^T \qquad (9)$$

and σ_1^2, Λ_2, and Λ_3 are the (co)variances of n_1, $\mathbf{n}_2$, and $\mathbf{n}_3$ respectively.

3.2. The posterior distribution of $\Delta\mathbf{p}$

We assume that a prior for $\Delta\mathbf{p}$ is available with an expected value of μ_r, and a covariance of Λ_r:

$$p(\Delta\mathbf{p}) = \frac{1}{\sqrt{|2\pi\Lambda_r|}} e^{-\frac{1}{2}(\Delta\mathbf{p} - \mu_r)^T \Lambda_r^{-1} (\Delta\mathbf{p} - \mu_r)} \qquad (10)$$

Prior information should improve the stability of image alignment as it represents something to balance against evidence from the pixel intensity difference measurements. Combining the intensity difference measurement model at a point (Equation 8) with the prior definition (Equation 10) using

Bayes rule (Equation 6) gives a Gaussian posterior estimate of $\Delta\mathbf{p}$ from the measurement at a single pixel of $\hat{\Delta\mathbf{p}}_s$ with a covariance of $\Lambda_{\Delta\mathbf{p},s}$.

$$\hat{\Delta\mathbf{p}}_s = \Lambda_{\Delta\mathbf{p},s}\left[\Lambda_r^{-1}\mu_r + \sigma_z^{-2}\left(\nabla T\frac{\partial \mathtt{W}}{\partial \mathbf{p}}\right)^T(I(\mathtt{W}(\mathbf{x};\mathbf{p})) - T(\mathbf{x}))\right]$$

$$\Lambda_{\Delta\mathbf{p},s} = \left[\Lambda_r^{-1} + \sigma_z^{-2}\left(\nabla T\frac{\partial \mathtt{W}}{\partial \mathbf{p}}\right)^T\left(\nabla T\frac{\partial \mathtt{W}}{\partial \mathbf{p}}\right)\right]^{-1}$$

Combining the measurements at individual pixels into an overall estimate of $\Delta\mathbf{p}$, assuming that the measurements are uncorrelated and noting that the prior is the same for all pixel measurements gives an estimate for $\Delta\mathbf{p}$ of:

$$\hat{\Delta\mathbf{p}} = \Lambda_{\Delta\mathbf{p}}\left[\Lambda_r^{-1}\mu_r + \sum_i \sigma_z^{-2}\left(\nabla T\frac{\partial \mathtt{W}}{\partial \mathbf{p}}\right)^T(I(\mathtt{W}(\mathbf{x};\mathbf{p})) - T(\mathbf{x}))\right] \quad (11)$$

$$\Lambda_{\Delta\mathbf{p}} = \left[\Lambda_r^{-1} + \sum_i \sigma_z^{-2}\left(\nabla T\frac{\partial \mathtt{W}}{\partial \mathbf{p}}\right)^T\left(\nabla T\frac{\partial \mathtt{W}}{\partial \mathbf{p}}\right)\right]^{-1} \quad (12)$$

This calculation is described in detail in [10].

3.3. Implementation

Equations 11, 12, and 9 are used in the following way. When each new image arrives for alignment, any available information about the warp is used to set the priors as follows:

- Set $\mathbf{p}$ to the prior estimate of the warp
- Set Λ_r to the prior estimate covariance of the warp
- Set μ_r to zero

Then, as the inverse compositional method runs, the value μ_r is set at each iteration to the parameter vector which represents the difference between the current warp estimate ($\mathbf{p}$) and the initial estimate of $\mathbf{p}$, set solely from the prior information. This ensures that during the iterative loop, where $\mathbf{p}$ is being updated, the prior is always effectively set to the starting $\mathbf{p}$ value. Iteration continues until the rms value of pixel intensity difference, $I(\mathtt{W}(\mathbf{x};\mathbf{p})) - T(\mathbf{x})$, drops below a threshold.

We applied Gaussian image smoothing with a standard deviation of one pixel to the images for the calculation of intensity gradient ∇T. The images were otherwise unsmoothed.

4. PROBABILISTIC ALIGNMENT EXAMPLE: AN AFFINE PATCH TRACKER

The affine warp has six degrees of freedom and can be written as:

$$\mathbf{x}' = \mathtt{A}\mathbf{x} + \mathbf{t}_a,$$

where $\mathbf{x}$, $\mathbf{x}'$, and $\mathbf{t}_a$ are 2-vectors and $\mathtt{A}$ is a 2x2 matrix. This is the simplest parameterisation of the affine warp in terms of six variables ($\mathtt{W}_s$) but it is difficult to assign prior information to these parameters during tracking. The natural interpretation of an affine transformation is as a translation of the patch centre (t_x, t_y), a rotation (by α), an isotropic scaling (s_i) and a non-isotropic scaling (s_n) in a particular direction (angle θ). This parameterisation is not suitable for use in the inverse compositional algorithm directly, because changing the value of θ has no effect when s_n is zero, giving a zero column in the Jacobian, but it can be modified to a form which is suitable while remaining intuitive and easy to apply prior estimates to:

$$\mathtt{A} = \frac{1+s_i}{s_t}\mathtt{R}_\alpha\begin{bmatrix} s_t^2\cos^2\theta + \sin^2\theta & (1-s_t^2)\cos\theta\sin\theta \\ (1-s_t^2)\cos\theta\sin\theta & s_t^2\sin^2\theta + \cos^2\theta \end{bmatrix}$$

$$\mathbf{t}_a = \mathbf{x}_0 - \mathtt{A}\mathbf{x}_0 + \mathbf{t},$$

$$\text{where,}\quad s_t = 1 + \sqrt{s_0^2 + s_{45}^2};\quad \mathtt{R}_\alpha = \begin{bmatrix}\cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha\end{bmatrix},$$

$$\cos\theta\sin\theta = \frac{1}{2}\left[\frac{s_{45}}{\sqrt{s_0^2+s_{45}^2}}\right];\quad \sin^2\theta = \frac{1}{2}\left[1 - \frac{s_0}{\sqrt{s_0^2+s_{45}^2}}\right]$$

$$\cos^2\theta = \frac{1}{2}\left[\frac{s_{45}^2}{s_0^2 + s_{45}^2 - s_0\sqrt{s_0^2+s_{45}^2}}\right].$$

and $\mathbf{x}_0$ is the centre position of the image patch in the template image $\mathtt{T}$. The $\mathbf{x}_0$ terms in the function for $\mathbf{t}_a$ ensure that a change in $\mathtt{A}$ does not change the position of the point $\mathbf{x}_0$ (the centre of the patch), so that the only way to translate the patch is through the vector $\mathbf{t}$. The effect of the parameter s_0 is to stretch along the x or y axis, and the parameter s_{45} stretches along the axes at 45 degrees to the x and y axes. These terms replace θ and s_n.

4.1. Live affine tracking tests

We tested the affine trackers firstly by running each tracker in real time on a sequence of images coming from a firewire camera. We compared three tracking methods: an affine version of the standard inverse compositional algorithm as described by Baker and Mathews; a reparameterised affine tracker using the parameterisation described above; and a probabilistic version of the parameterised tracker.

From these tests it was apparent that the parameterised trackers were both significantly more stable than the standard affine tracker. This is particularly noticeable for image regions which are invariant to some components of the affine transformation. For example a texture with one homogenous colour in the bottom left quadrant and a different homogenous colour elsewhere (such as might be seen when viewing the corner of an object) does not change in appearance under changes in scale. Ideally we would write

a feature detector which scores image regions based on variability under each component of the alignment warp, to ensure that we are always tracking features for which all components of the warp are observable (The Shi Tomasi detector [3] is designed to do this but for image translation only). The problem is limited to a certain extent when using a probabilistic tracker, which, because it places a prior on each component of the warp, at least remains stable when tracking this type of texture.

4.2. Rendered image sequence tests

We also ran a series of partially synthetic tests for a more quantitative assessment of performance. A number of synthetic image sequences were generated by loading an image from disk and projecting it into a virtual camera moving along a path in 3D. The image sequences were generated by bilinear interpolation of the file image, and adding noise to each pixel, uniformly distributed between plus and minus five intensity values.

We used up to 10 different images and 10 different synthetic camera motions, each sequence being between 100 and 200 frames long. The camera moves were designed to reproduce common motions which might happen when a camera moves around a scene.

We ran the Shi Tomasi feature detector [3] to find the best 50 features in the first image of the sequence and then eliminated any features corresponding to positions on the plane which would not be in view for the whole of the remainder of the sequence. This typically resulted in 20-40 trackers per sequence, each running on a 31 pixel square image patch. Across all sequences, the corners of the tracked patches moved on average by 2.52 pixels between each image, although for some sequences the variation was significantly more or less than this. Tracking of each feature was carried out by performing alignment at each frame, using the alignment result from the previous frame as a starting point. This test highlights differences in performance, since a failure in alignment at any point makes it vary hard to maintain tracking for the remainder of the sequence.

4.3. Rendered image sequence results

We evaluated the trackers by defining an error measure for each alignment as the average difference in image position of the four corners of the image patch when warped by the calculated affinity and by the ground truth homography. Because the affine tracker cannot reproduce the eight degrees of freedom of the homography, the two transformations will rarely line up exactly, but a better alignment algorithm will give a closer result.

Figure 2 compares the distribution of final alignment error for the three alignment methods using the full set of 100 synthetic image sequences. Alignments giving errors

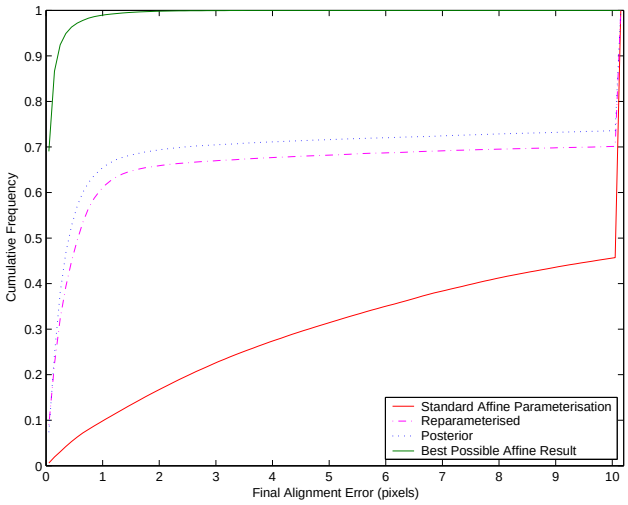

Fig. 2. Cumulative distribution of alignment error for the three trackers

above 10 pixels are collected in the last data points in the graph. The value of each graph at a particular alignment error value indicates the proportion of alignments made to within this level of error. The graphs show that the best results consistently come from the fully probabilistic tracker, followed by the parameterised tracker, and lastly the standard affine tracker. In terms of speed the ranking is reversed, with the standard affine alignment method taking on average 0.531ms, the reparameterised method 0.658ms, and the probabilistic method 0.775ms. These timing values are however biased towards trackers which are badly initialised and therefore stop iterating after the first few iterations, which have failed to improve the alignment score.

The improved results following re-parameterisation may be due to the fact that the new parameterisation relates different types of image change more closely to parameters. For example, patch translation, which is usually the greatest cause of intensity change, can occur by changing many different combinations of the six parameters of the standard affine model, but is isolated to two parameters in the reparameterised model. The intensity error surface in parameter space might also be more stable after the re-parameterisation, or the range of motions spanned by small changes around the origin of parameter space might be more limited, resulting in greater stability.

The use of prior information during alignment was clearly of benefit, although choosing the prior values is not always straightforward. We chose prior uncertainty values which were large enough to cover the large range of camera motions in the test set. With more certain prior information, from, for example a Kalman Filter tracking groups of image features through time, the probabilistic alignment method should give stronger results.

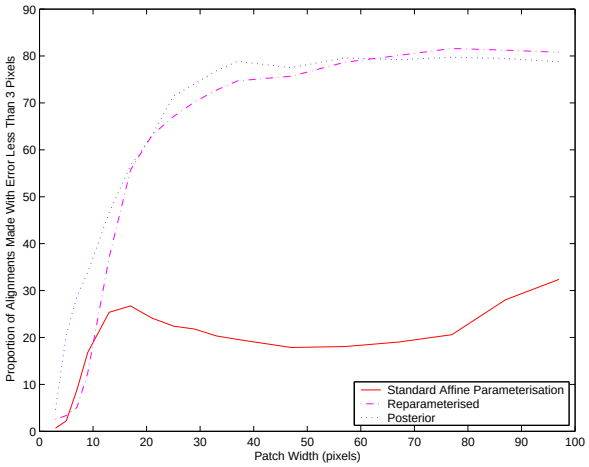

Fig. 3. The effect of changing patch size on the number of alignments made with an error less than 2 pixels

We also looked briefly at the effect of changing the patch size used by the trackers on the proportion of alignments resulting in an error of less than 3 pixels (shown in Figure 3) and on tracking time. The performance advantage of the probabilistic tracker is greatest when the patch size is relatively small. As the patch size increases, the number of pixel measurements made rises and the relative significance of the prior falls, so that eventually the probabilistic tracker performs very similarly to the parameterisaed tracker. As the number of pixels in the patch increases, the average alignment time increases linearly, as would be expected, and the difference in alignment time between the methods remains fixed. Besides speed issues, a smaller patch size is also desirable because the assumption that parts of the image are locally planar becomes more credible.

5. CONCLUSIONS

We have looked at improving the stability of the inverse compositional image alignment algorithm by incorporating a physically motivated warp model, prior uncertainty, and image intensity measurement uncertainty in a probabilistic estimation process. Like the original inverse compositional method, the methods we describe can be run and initialised in real time. Further detail of the work described here can be found in [10].

Changing the parameterisation of the warp function can be done quite straightforwardly, with little computational cost, though care must be taken to ensure that the new warping function does not introduce excessive nonlinearities. Reparameterisation can be useful in its own right, and this is demonstrated in Section 4 where a non-traditional parameterisation of the affine transformation greatly improved the stability of affine tracking, even without the probabilistic methods described in the rest of this paper.

The probabilistic version of image alignment, with priors set appropriately gave an improvement in alignment stability for a modest increase in running time. By changing the prior uncertainty the amount of freedom given to each parameter can be changed at will and even set to zero, removing the parameter from the optimisation. This is significant because the basic inverse compositional method is very sensitive to the freedom in the warp parameters.

Although we looked at re-parameterisation and probability for the inverse compositional image alignment method here, the logic would be very similar for comparable gradient-based methods, such as the original Lucas and Kanade tracker. The ability to change the parameterisation at will and to use uncertainty information demonstrates that these methods may be stable and flexible enough to be applied to a wider range of applications than they have been used in to date, for example, in tracking textured non-rigid objects.

6. REFERENCES

[1] S. Baker and Iain Matthews, "Lucas-Kanade 20 years on: A unifying framework: Part 1," Report CMU-RI-TR-02-16, Carnegie Mellon University, 2002.

[2] B.D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," in *IJCAI81*, 1981, pp. 674–679.

[3] Jianbo Shi and Carlo Tomasi, "Good features to track," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR'94)*, Seattle, June 1994.

[4] T. Tommasini, A. Fusiello, E. Trucco, and V. Roberto, "Making good features track better," in *Proc. Computer Vision and Pattern Recognition*, Santa Barbara, California, USA, June 1998, pp. 178–183, IEEE Computer Society Press.

[5] H. Jin, P. Favaro, and S. Soatto, "Real-Time feature tracking and outlier rejection with changes in illumination," in *ICCV*, July 2001, pp. 684–689.

[6] S. Baker, Ralph Gross, Takahiro Ishikawa, and Iain Matthews, "Lucas-Kanade 20 years on: A unifying framework: Part 2," Report CMU-RI-TR-03-01, Carnegie Mellon University, 2003.

[7] J.L. Barron, D.J. Fleet, S.S. Beauchemin, and T.A. Burkitt, "Performance of optical fbw techniques," *CVPR*, vol. 92, pp. 236–242.

[8] F. Jurie and M. Dhome, "Real time robust template matching," in *British Machine Vision Conference 2002*, Cardiff, 2002, pp. 123–131.

[9] R. Hartley and A. Zisserman, *Multiple View Geometry in Comuter Vision*, Cambridge University Press, 2000.

[10] N. D. Molton, A. Davison, and I. Reid, "Parameterisation and probability in image alignment," Report OUEL 2266/03, Department of Engineering Science, University of Oxford (Available from http:/www.robots.ox.ac.uk/˜ndm/), Aug. 2003.