Generalizing inverse compositional image alignment

Rupert Brooks and Tal Arbel
McGill University
Montreal, Quebec, Canada
{rupert.brooks,tal.arbel}@mcgill.ca

Abstract

The inverse compositional (IC) approach to image alignment uses characteristics of the alignment problem to improve optimization speed. While a number of authors have noted its usefulness, to date it has only been explored for least-squares type image difference measures using Gauss-Newton optimization schemes. We extend the IC approach to general difference measures, and a wider class of optimization approaches, with specific development for normalized correlation and mutual information using the BFGS optimizer. We present alignment experiments on image pairs of several different classes that demonstrate performance improvements for the general case.

1. Introduction

Image alignment, or registration, is one of the oldest and most widely used computer vision techniques [5, 11]. Certain applications, such as real-time tracking or image guided surgery require computationally efficient approaches. Recently, Baker and Matthews introduced inverse compositional (IC) image alignment [3]. This algorithm consists of two modifications to the Lucas-Kanade algorithm [8] which lead to greater computational efficiency in a number of reported contexts [10, 1]. However, to date it has only been applied to problems where the image difference measure is a sum of squares, which has in turn lead to it being used almost exclusively with Gauss-Newton optimization.

In this paper, we show that the advantages of this approach apply more generally than previously described. In section 2 we review direct image alignment, and explain the advantages of the IC method. In section 3 we describe how the approach may be extended to any image difference measure, and any first-order optimization technique. In section 4 we describe an implementation of the generalized approach for the normalized correlation (NC) and mutual information (MI) measures using a Broyden-Fletcher-Goldfarb-Shanno (BFGS) optimizer. Finally, in section 5, we show the performance improvements that result.

2. Image Alignment

2.1. Direct image alignment

Given two images, $R(\bar{x})$ and $T(\bar{x})$, the alignment problem is to find a spatial transformation, or warp, $W(\bar{x};\phi)$, (parameterized by ϕ) which will map the moving image $T(W(\bar{x};\phi))$ onto corresponding parts of the fixed reference image, $R(\bar{x})$. There are two major approaches to the problem [5, 11]. *Direct* approaches define a difference measure, D, between the reference image and the warped template image. The solution is the warp which minimizes this difference, and is generally found using numerical optimization techniques. In contrast, feature based approaches address the problem by finding invariant features in each image and matching them. The matched set of features can then be used to determine the warp. While feature based approaches have seen a number of recent successes, direct image alignment can yield higher overall accuracy and works in cases where feature based approaches break down, such as on multi-modal imagery. Not surprisingly, direct approaches are widely used in geomatics and medical imaging where multi-modal imagery is common and local accuracy is important [5, 11].

In this paper we address direct image alignment which can be expressed as a standard optimization problem:

$$\phi_{opt} = \underset{\phi}{\operatorname{argmin}} \left(D(R(\bar{x}), T(W(\bar{x};\phi))) \right), \qquad (1)$$

where ϕ_{opt} are the warp parameters that minimize the image difference D. Gradient based optimization techniques are frequently used to find ϕ_{opt}. These methods iteratively update an estimate of the optimal parameters, by computing a parameter update $\Delta\phi$ and adding it to the current parameter estimate ϕ_i, until some convergence criteria are met.

2.2. Inverse Compositional method

Baker and Matthews [3] propose the inverse compositional (IC) method as a faster algorithm for image alignment. They consider the direct image alignment method

The 18th International Conference on Pattern Recognition (ICPR'06)
0-7695-2521-0/06 $20.00 © 2006 IEEE

developed by Lucas and Kanade [8], which uses the sum of squared differences of corresponding image intensities as a difference measure and optimizes it using a Gauss-Newton technique. Baker and Matthews argue that making a small change to the warp parameters of the moving image is equivalent to warping the reference image slightly in the opposite direction. Thus they use an *inverse* approach - instead of searching for a parameter update which would better warp the moving image to the reference image, they search for a warp update, $\Delta\psi$, which would warp the fixed image to the current moving image and then update the parameters with the inverse of that warp. To be valid, the parameter update has to be *compositional*, that is the update step becomes $\phi_{i+1} = \phi_i \circ \Delta\psi^{-1}$, where $\circ$ is the warp composition operator. This method theoretically requires that the warps form a group, but in practice it seems to work so long as the warps behave like a group in the region of interest [9].

This method is significantly faster because the search is always performed around the zero warp of the fixed image, and therefore the image gradient, the Jacobian and the Gauss-Newton approximation to the Hessian may be precomputed. This idea has been applied in several contexts, including active appearance models [9], and 3D medical data [1, 4]. However, in all cases, the similarity measure used has been the sum of (possibly robustly reweighted) squared pixel intensity differences, and the optimization methods have relied on the structure of this similarity measure rather than being general.

3. Generalizing the IC Approach

In [3] there is a comparison of optimization methods, which demonstrate that the Gauss-Newton approach is superior for a sum of squared differences measure. However, this optimization method is only suitable for difference measures of that form, and the variable metric methods generally considered as the superior methods for general multivariable optimization [13] are not addressed.

The computational advantages of the inverse compositional approach, however, are not limited to measures made of summed squared residuals. In fact, this method can be applied to the optimization of any function of the form:

$$\phi_{opt} = \underset{\phi}{\operatorname{argmin}} \left(D(R(W(\bar{x};\psi)), T(W(\bar{x};\phi))) \right), \qquad (2)$$

where a change in the warp parameters of the moving image, $\Delta\phi$, is equivalent to changing the warp of the fixed image by $\Delta\psi = \Delta\phi^{-1}$, provided the restrictions in [3] hold.

3.1. General IC Optimization

Gradient based numerical optimization techniques (Algorithm 1) find the minimum of an objective function, $F(\phi)$, by iteratively updating the parameters, ϕ, until some convergence criteria are reached. Algorithms primarily differ in how the update, $\Delta\phi$, is computed.

Algorithm 1 Gradient-based Optimization Algorithm

Set start position ϕ_0; iteration counter $i = 0$
repeat
  Compute update $\Delta\phi$ based on $F(\phi_i + \Delta\phi), \nabla F_\phi|_{\phi=\phi_i}$
  Set $\phi_{i+1} = \phi_i + \Delta\phi$ and $i = i + 1$
until convergence criteria reached

To extend the IC approach to general gradient based optimization algorithms, note that this approach uses the objective function gradient with respect to ψ rather than ϕ. This forces the optimization algorithm to search a different parameter space for an update step $\Delta\psi$. This step corresponds to an update $\phi_{i+1} = \phi_i \circ \Delta\psi^{-1}$ in the original parameter space. Therefore, at each iteration we create a new objective function, $G_i(\psi) = F(\phi_i \circ (\psi - \psi_i)^{-1})$, that is provided to the optimization algorithm. Note that this is not simply a reparameterization. $G_i(\psi)$ is dependent on ψ_i and therefore must be updated at each iteration, so that the additive changes in ψ properly correspond to compositional changes in ϕ. This is shown in algorithm 2.

Algorithm 2 General Inverse Compositional Optimization

Set start position ϕ_0; iteration counter $i = 0$
Set $\psi_0 = \phi_0$
repeat
  Define: $G_i(\psi) = F(\phi_i \circ (\psi - \psi_i)^{-1})$
  Define: $\nabla G_i(\psi) = \nabla_\psi F(\phi_i \circ (\psi - \psi_i)^{-1})$
  Compute update $\Delta\psi$ based on $G_i(\psi)$ and $\nabla_\psi G_i(\psi)$
  Set $\phi_{i+1} = \phi_i \circ \Delta\psi^{-1}$ and $i = i + 1$
  Set $\psi_{i+1} = \psi_i + \Delta\psi$
until convergence criteria reached

4. Implementation

4.1. Image difference measures

Two image difference measures in wide use are the *normalized correlation* (NC) measure [2] and the mutual information (MI) measure [14]. Normalized correlation is appropriate for comparing images where there is a linear relationship between the intensities in the fixed and moving images, such as photographs under different lighting conditions or satellite images in different spectral bands. Expressed in terms of the variables in equation 2, the NC measure is given by equation 3 and its gradient with respect to ψ by equation 4. While NC relies on the image intensities being correlated when they are in proper alignment, MI is more

$$D_{NC}(\phi) = \frac{\sum_{\bar{x}} \left[(R(W(\bar{x};\psi)) - \bar{R}(W(\bar{x};\psi)))(T(W(\bar{x};\phi)) - \bar{T}(W(\bar{x};\phi))) \right]}{\sqrt{\left[\sum_{\bar{x}} (R(W(\bar{x};\psi)) - \bar{R}(W(\bar{x};\psi)))^2\right] \left[\sum_{\bar{x}} (T(W(\bar{x};\phi)) - \bar{T}(W(\bar{x};\phi)))^2\right]}} = \frac{U}{V} \quad (3)$$

$$\nabla_\psi D_{NC}(\phi) = \frac{\sum_{\bar{x}} \left[\nabla_\psi R(W(\bar{x};\psi))(T(W(\bar{x};\phi)) - \bar{T}(W(\bar{x};\phi))) \right] - U \cdot \frac{\sum_{\bar{x}} \left[\nabla_\psi R(W(\bar{x};\psi))(R(W(\bar{x};\psi)) - \bar{R}(W(\bar{x};\psi))) \right]}{\sum_{\bar{x}} \left[(R(W(\bar{x};\psi)) - \bar{R}(W(\bar{x};\psi)))^2 \right]}}{V} \quad (4)$$

$$D_{MI}(\phi) = -\sum_{\iota,\kappa} P_{\iota,\kappa}(\phi) \log \frac{P_{\iota,\kappa}(\phi)}{\sum_\kappa P_{\iota,\kappa}(\phi) \sum_\iota P_{\iota,\kappa}(\phi)} \quad (5)$$

$$\nabla_\psi D_{MI}(\phi) = \sum_{\iota,\kappa} \frac{\partial P_{\iota,\kappa}(\phi)}{\partial \psi}(\phi) \log \frac{P_{\iota,\kappa}(\phi)}{\sum_\iota P_{\iota,\kappa}(\phi)} \quad (6)$$

general, and reveals the degree to which the two signals are not independent. Thus it is appropriate for cases where the imaging modalities are very different and the relationship between the image intensities is complex. Implementation details are very important for achieving good results with this measure and a direct implementation is rarely effective. In this work we use the efficient formulation in [12], which computes the MI based on a model of the joint intensity distribution $P_{\iota,\kappa}(\phi)$ and its derivative $\frac{\partial P_{\iota,\kappa}(\phi)}{\partial \phi}$. The reader is referred to [12] for the details. In the original formulation, the gradient with respect to ψ does not exist. However, we modify the technique to compute the gradient with respect to ψ by simply reversing the roles of the reference and the fixed image for the gradient part of the calculation only. Then, the MI and its gradient may be computed using equations 5 and 6, respectively.

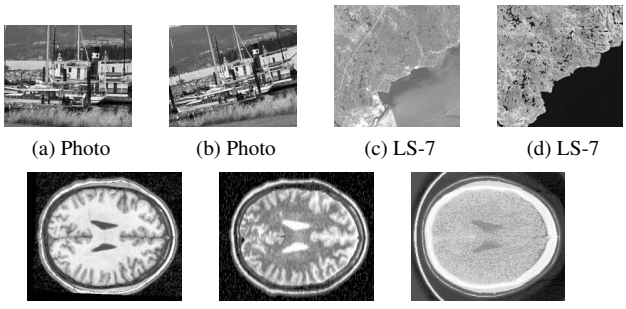
4.2. Optimizers

The BFGS optimization method is considered to be the superior method for general first order numerical optimization [13]. Briefly, it performs its update step by selecting a search direction in the parameter space, and then performing a line search in the direction of interest. At each step it updates an internal estimate of the Hessian. Over time it is able to use this estimate to search in directions which are conjugate to the Hessian, leading to rapid convergence.

In this work, we extended the BFGS implementation described by [6] to support IC alignment. This implementation requests evaluations of the objective function (and its gradient) at various parameter values. Thus we were able to implement our approach by maintaining the relationship between the optimizer space, ψ, and the parameter space ϕ and feeding the algorithm values of D and $\nabla_\psi D$ evaluated at $\phi_i \circ \Delta\psi^{-1}$ for each requested evaluation at ψ.

5. Experiments

The ITK framework [7] provides implementations of many standard image alignment algorithms and optimizers, including the NCC and MI measures and the BFGS optimizer described in [6]. We extended the existing implementation to support the IC versions of NC and MI and the optimizer. The algorithms were tested by aligning image pairs of different classes (shown in Figure 1) for which gold standard transformations were independently available using both the original and IC implementations. Each pair was aligned using both difference measures and in both directions, that is, each image in the pair was used as both the fixed image and the moving image. Image pair a-b was aligned using an affine transformation, while the others were aligned using rigid transformations. In every case the alignment process was started from 60 different positions ranging from relatively close to considerably misaligned. The results of all these tests are shown in Table 1.



(a) Photo (b) Photo (c) LS-7 (d) LS-7

(e) T1 MRI (f) T2 MRI (g) CT

Figure 1: Image Pairs

Images a, b are photographs taken from different camera positions. Images c, d are co-registered Landsat 7 images taken at different times in different spectral bands. Band 3 (c) is 630-690 nm (red) and band 4 (d) is 780-900 nm (IR). Images e-g are co-registered slices from the same brain image in different modalities.[1]

For each dataset, the performance of the algorithms was measured using its CPU time (in seconds), RMS accuracy (in pixels) and failure rate. The CPU time refers to runs using an implementation entirely in C++, running under Linux on a 3.06GHz Pentium 4. RMS accuracy was measured by projecting a set of 50 points scaled to the image extent through the found transform, and the known, gold standard

[1] Images a), b) from K. Mikolajczyk `http://www.inrialpes.fr/lear/people/Mikolajczyk/`; Images c), d) from Natural Resources Canada `http://geogratis.gc.ca`; Images e), f), g) courtesy Montreal Neurological Institute

transform. A run was considered a failure when either the algorithm failed to complete, or the final RMS error was more than five pixels.

	Pair	D	Alg	RMS	Fail	Time	ΔT
1	a-b	NC	STD	0.53	6.7%	29.65	
2	a-b	NC	IC	0.55	6.7%	20.14	32%
3	b-a	NC	STD	0.77	6.7%	26.69	
4	b-a	NC	IC	0.81	6.7%	20.24	24%
5	c-d	NC	STD	1.34	3.3%	197.40	
6	c-d	NC	IC	1.29	10.0%	38.62	80%
7	d-c	NC	STD	0.39	0.0%	106.46	
8	d-c	NC	IC	0.53	0.0%	97.52	8%
9	e-f	MI	STD	0.40	10.0%	9.96	
10	e-f	MI	IC	0.39	10.0%	5.68	42%
11	f-e	MI	STD	0.40	16.7%	9.98	
12	f-e	MI	IC	0.44	15.0%	5.29	46%
13	e-g	MI	STD	0.36	8.3%	13.34	
14	e-g	MI	IC	0.39	11.7%	5.83	56%
15	g-e	MI	STD	0.62	16.7%	10.27	
16	g-e	MI	IC	0.60	20.0%	5.90	42%
17	f-g	MI	STD	1.07	8.3%	11.89	
18	f-g	MI	IC	0.85	3.3%	6.39	46%
19	g-f	MI	STD	0.62	16.7%	10.08	
20	g-f	MI	IC	0.60	20.0%	5.79	42%

Table 1: Experimental Results

Overall, we found that the standard and IC approach are similar in terms of accuracy and robustness, with the IC approach being faster, which roughly confirms the results in [3]. However, our results differ from [3] in several important ways. We find that the speed up (ΔT) due to the IC method, is not an order of magnitude as they report, but averages about 42%. This is most likely because in [3] parts of the algorithm were implemented in C and parts in Matlab. A greater percentage of the IC algorithm was implemented in C which may have yielded better performance. We also found a greater variation in robustness between the two approaches. This is not surprising either, as in [3] the same image, prewarped using the same interpolator as the algorithm, is used as both the fixed and reference image. In our case, particularly with multi-modal data, the fixed and reference images are very different from each other, and the derivative of one is likely to be much noisier than the other. Thus we find, (e.g. rows 5-8 of Table 1) that, performance depends on which image modality is used as the fixed image. Which modality should be made fixed to get best computational performance will be reversed between the normal, and IC approaches.

6. Conclusions

We have shown that the computational advantages of the inverse compositional approach apply to a wider class of image alignment problems than have previously been addressed. We develop the approach for general image difference measures and general optimization techniques and demonstrate its effectiveness using the commonly used normalized correlation and mutual information measures.

Our results confirm previous work on more limited cases, showing that the inverse compositional method is as reliable as the standard method, but runs significantly faster. However, we do not see the same order of magnitude speedup, but a more modest, but still important speedup of about 42%. Furthermore, for multi-modal imagery the selection of which type of image should be made fixed and which should move is opposite when using the IC method.

References

[1] A. Andreopoulos and J. K. Tsotsos. A novel algorithm for fitting 3d active appearance models: Applications to cardiac mri segmentation. In *Proceedings of the 14th Scandinavian Conference on Image Analysis*, Joensuu, Finland, 2005.

[2] P. E. Anuta. Digital registration of multispectral imagery. *SPIE Journal*, 7(6):168–175, 1969.

[3] S. Baker and I. Matthews. Lucas-Kanade 20 years on: A unified framework. *Int. J. Comp. Vision*, 56(3):221–255, 2004.

[4] S. Baker et al. Lucas-Kanade 20 years on: Part 5. Technical Report CMU-RI-TR-04-64, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, 2004.

[5] L. G. Brown. A survey of image registration techniques. *ACM Computing Surveys*, 24(4):325–376, December 1992.

[6] R. H. Byrd, P. Lu, and J. Nocedal. A limited memory algorithm for bound constrained optimization. *SIAM J. Scientific and Statistical Comput.*, 16(5):1190–1208, 1995.

[7] L. Ibanez et al. *The ITK Software Guide: ITK V2.0*. Kitware Inc, 2005. `http://www.itk.org`.

[8] B. D. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *Proc. of Image Understanding Workshop*, pages 121–130, 1981.

[9] I. Matthews and S. Baker. Active appearance models revisited. *Int. J. Comp. Vision*, 60(2):135 – 164, 2004.

[10] N. D. Molton, A. J. Davison, and I. D. Reid. Parameterisation and probability in image alignment. In *Proc. Asian Conference on Computer Vision*, Jeju, Jan. 2004.

[11] R. Szeliski. Image alignment and stitching: A tutorial. Technical Report MSR-TR-2004-92, Microsoft Research, 2004.

[12] P. Thévenaz and M. Unser. Optimization of mutual information for multiresolution image registration. *IEEE Trans. Image Processing*, 9(12):2083–2099, 2000.

[13] P. Venkataraman. *Applied Optimization with MATLAB Programming*. John Wiley and Sons, Canada, 2001.

[14] P. Viola and W. M. Wells III. Alignment by maximization of mutual information. In *Proc. of the Fifth International Conference on Computer Vision,*, pages 16–23, 1995.