Efficient subpixel image registration algorithms

Manuel Guizar-Sicairos, Samuel T. Thurman, and James R. Fienup*

The Institute of Optics, University of Rochester, Rochester, New York, 14627, USA
**Corresponding author: fienup@optics.rochester.edu*

Received September 7, 2007; revised December 7, 2007; accepted December 7, 2007;
posted December 12, 2007 (Doc. ID 87296); published January 10, 2008

Three new algorithms for 2D translation image registration to within a small fraction of a pixel that use nonlinear optimization and matrix-multiply discrete Fourier transforms are compared. These algorithms can achieve registration with an accuracy equivalent to that of the conventional fast Fourier transform upsampling approach in a small fraction of the computation time and with greatly reduced memory requirements. Their accuracy and computation time are compared for the purpose of evaluating a translation-invariant error metric. © 2008 Optical Society of America

OCIS codes: 100.2000, 100.3020, 100.5070, 100.5010.

In a wide variety of applications it is often desired to register two images to within a small fraction of a pixel for image processing tasks or assessment. In this work we are primarily concerned with evaluation of reconstructed images by phase retrieval [1,2], for which a 2D rigid translation is sufficient. Other applications can also be found, though a more general transformation is often required, in processing of remotely sensed data, biological and medical imaging, and computer vision, with associated tasks ranging from superresolution [3], speckle and noise reduction [4], motion and change tracking, and stereoscopic vision [5], among others [6–8].

For the case of just a translation between two images, the usual technique to address this problem is to compute an upsampled cross correlation between the image to register and a reference image, by means of a fast Fourier transform (FFT), and locating its peak. The computational burden associated with such an approach increases as the required accuracy of the registration increases, especially in terms of memory. For example, registration to within 1/20 of a pixel for 1024×1024 images requires computation and storage of a $20{,}480\times20{,}480$ inverse FFT, which cannot be done currently on most personal computers.

In this Letter we present three efficient registration algorithms based on nonlinear optimization and discrete Fourier transforms (DFTs) that allow accurate registration of two images with large upsampling factors. We test the performance of these algorithms, in terms of accuracy and computational speed, when used in the computation of a translation-invariant error metric for image quality. When compared with the usual FFT approach, these algorithms have much shorter computation times and greatly reduced memory requirements.

For the problem of image reconstruction by phase retrieval, an image $g(x,y)$ of an object $f(x,y)$, can be reconstructed numerically from measurements of the magnitude of the Fourier transform of $f(x,y)$ [1,2]. In this context, a reconstruction $g(x,y)$ is considered successful even if it has a global coordinate translation (x_0,y_0) or is multiplied by an arbitrary constant α. The quality of the reconstruction must then be assessed through an error metric that is invariant to these operations. One such metric is the normalized root-mean-square error (NRMSE) E between $f(x,y)$ and $g(x,y)$, defined by [9]

$$E^2 = \min_{\alpha,x_0,y_0} \frac{\sum_{x,y} |\alpha g(x-x_0,y-y_0) - f(x,y)|^2}{\sum_{x,y} |f(x,y)|^2} = 1 - \frac{\max_{x_0,y_0} |r_{fg}(x_0,y_0)|^2}{\sum_{x,y} |f(x,y)|^2 \sum_{x,y} |g(x,y)|^2}, \tag{1}$$

where summations are taken over all image points (x,y);

$$r_{fg}(x_0,y_0) = \sum_{x,y} f(x,y) g^*(x-x_0,y-y_0) = \sum_{u,v} F(u,v) G^*(u,v) \exp\left[i2\pi \left(\frac{ux_0}{M} + \frac{vy_0}{N} \right) \right]; \tag{2}$$

is the cross correlation of $f(x,y)$ and $g(x,y)$; N and M are the image dimensions; $(*)$ denotes complex conjugation; uppercase letters represent the DFT of their lowercase counterparts, as given by the relation

$$F(u,v) = \sum_{x,y} \frac{f(x,y)}{\sqrt{MN}} \exp\left[-i2\pi \left(\frac{ux}{M} + \frac{vy}{N} \right) \right]; \tag{3}$$

and E^2 is minimized with respect to α [9].

Thus, evaluation of the NRMSE by Eq. (1) requires solving the more general problem of subpixel image registration by locating the peak of the cross correlation $r_{fg}(x,y)$.

The usual FFT approach to finding the cross-correlation peak to within a fraction, $1/\kappa$, of a pixel is (i) compute $F(u,v)$ and $G(u,v)$, (ii) embed the product $F(u,v)G^*(u,v)$ in a larger array of zeros of dimensions $(\kappa M, \kappa N)$, (iii) compute an inverse FFT to obtain an upsampled cross correlation, and (iv) locate its peak. The computational complexity of the inverse FFT in this case is $O\{MN\kappa[\log_2(\kappa M)+\kappa\log_2(\kappa N)]\}$ for $N\le M$.

0146-9592/08/020156-3/$15.00 © 2008 Optical Society of America

As efficient alternatives to this approach, we developed three different algorithms that significantly improve performance without sacrificing accuracy. All three algorithms start with an initial estimate of the location of the cross-correlation peak obtained by the FFT method with an upsampling factor of $\kappa_0=2$. This initial upsampling is used in an effort to select an appropriate starting point for cross correlations that might have more than one peak of similar magnitude.

The first algorithm, first suggested in [9] but not implemented, refines the initial estimate using a nonlinear-optimization conjugate-gradient routine [10] to maximize $|r_{fg}(x_0,y_0)|^2$. Its partial derivative with respect to x_0 is

$$\frac{\partial|r_{fg}(x_0,y_0)|^2}{\partial x_0} = 2\,\mathrm{Im}\left\{ r_{fg}(x_0,y_0)\sum_{u,v}\frac{2\pi u}{M}F^*(u,v) \times G(u,v)\exp\left[-i2\pi\left(\frac{ux_0}{M}+\frac{vy_0}{N}\right)\right]\right\}, \tag{4}$$

with a similar expression for the partial derivative with respect to y_0. This algorithm iteratively searches for the image displacement (x_0,y_0) that maximizes $r_{fg}(x_0,y_0)$ and can achieve registration precision to within an arbitrary fraction of a pixel.

The second algorithm, which we will refer to as the single-step DFT approach, uses a matrix multiplication implementation of the 2D DFT [11,12], Eq. (3), to refine the initial peak location estimate. A 2D FFT is the most efficient approach when computation of all points of the upsampled cross correlation is required. Unfortunately the FFT is restricted to computing the entire upsampled array, of dimensions $(\kappa M,\kappa N)$, resulting in an enormous waste of resources if we are interested only in computing an upsampled version of $r_{fg}(x_0,y_0)$ in a very small neighborhood about the initial estimate of the peak location. The advantage of a matrix-multiply DFT results from the fact that an upsampled version of $r_{fg}(x_0,y_0)$ can be computed within just such a neighborhood without the need to zero-pad $F(u,v)G^*(u,v)$. In the single-step DFT algorithm, an upsampled cross correlation (by a factor κ) is computed in a 1.5×1.5 pixel neighborhood (in units of the original pixels) about the initial estimate. This operation is implemented by the product of three matrices with dimensions $(1.5\kappa,N)$, (N,M), and $(M,1.5\kappa)$. Subpixel registration is achieved by searching for the peak in the output $(1.5\kappa,1.5\kappa)$ array (in units of upsampled pixels). Assuming that for cases of interest κ is smaller than M and N, the algorithm complexity for this upsampling is $O(MN\kappa)$; a substantial improvement over the FFT upsampling approach.

As the required registration accuracy is increased, it proves useful to further reduce the amount of computed samples by taking a two-step matrix-multiply DFT approach when refining the initial $(\kappa_0=2)$ translation estimate. The two-step DFT algorithm initially upsamples a 1.5×1.5 pixel region by a factor $\kappa_1\simeq\kappa^{1/2}$ about the initial estimate and finds the cross-correlation peak in that array. In the second step the peak location is further refined by upsampling a $3/\kappa_1\times3/\kappa_1$ smaller region of the original pixel grid around the new estimate by a full factor of κ. In this manner the complexity of the refinement algorithm is reduced to $O(MN\kappa^{1/2})$. The precision, κ^{-1}, for this translation estimation is the same for the FFT, single- and two-step DFT approaches.

To assess and compare the performance of each algorithm, a 256×256 complex-valued image $f(x,y)$ was corrupted by additive zero-mean circular complex Gaussian noise $n(x,y)$ and translated by $(x_0,y_0)=(502/21,52/15)$ pixels to create $g(x,y)$. Results of the estimation of the invariant NRMSE, $\hat{E}$, in this case are shown in Fig. 1(a) with respect to the true NRMSE E, where $E^2=\Sigma_{x,y}|n(x,y)|^2/\Sigma_{x,y}|f(x,y)|^2$. As the NRMSE approaches zero, the use of high upsampling factors κ or the nonlinear optimization routine becomes crucial to avoid overestimation of E. The accuracy of the estimated translations $(\hat{x}_0,\hat{y}_0)$ was determined by calculating the shift error $\Delta r=[(x_0-\hat{x}_0)^2+(y_0-\hat{y}_0)]^{1/2}$, shown versus κ in Fig. 1(b) for $E=0.25$. For comparison purposes, Δr obtained with the nonlinear optimization routine, which does not require specification of κ, is shown as well. For this large value of $E=0.25$, the estimation $\hat{E}$ had diminishing returns from using upsampling factors higher than $\kappa=10$ [Fig. 1(a)], but it is clear from Fig. 1(b) that the registration accuracy benefits significantly from larger values of κ and even more from the use of the nonlinear optimization algorithm. Ultimately the registration accuracy will be limited by noise.

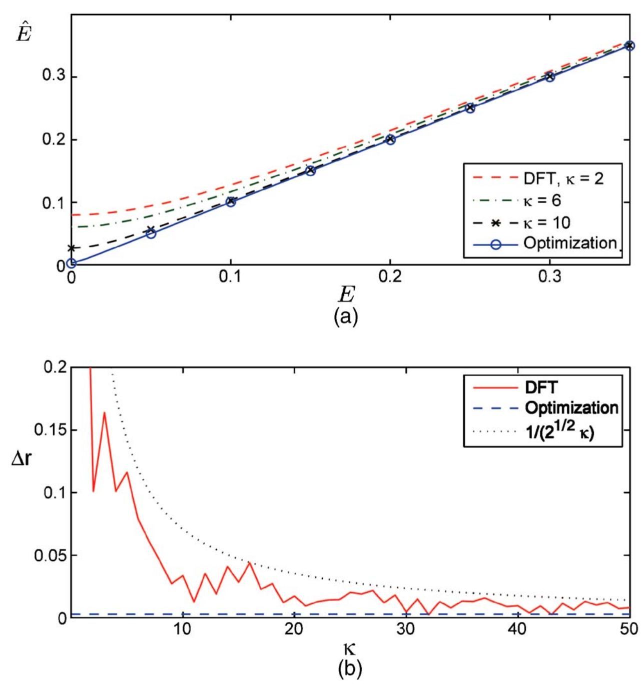

Fig. 1. (Color online) (a) Invariant NRMSE estimation $\hat{E}$ versus E. (b) Error in estimated image shift Δr versus upsampling factor κ for $E=0.25$. The dotted curve shows the noiseless maximum error, $1/(\sqrt{2}\kappa)$. For the optimization algorithm, $\Delta r=0.0029$ pixels (dashed curve).

A comparison of the computation time required for each of the three algorithms as obtained on a desktop computer (AMD Athlon X2 dual core processor 2.21 GHz, 64-bit operating system, 4 Gbytes RAM) is shown in Fig. 2. The time required to obtain the initial estimate by the $2\times$ upsampled FFT approach was included in the time measurement and explicitly shown for comparison purposes. The initial estimate was found to dominate the total computation time. Figure 2(a) shows the computation time for the registration algorithms with respect to image size for $E=0.22$ (and $\kappa=25$ for the DFT algorithms). The computation time as a function of κ for 512×512 images with the same amount of noise is shown in Fig. 2(b). The computation time of the nonlinear optimization algorithm is shown for comparison; it is independent of κ. The largest array size for which an FFT could be performed on the same computer is $11{,}585\times 11{,}585$, assuming two double-precision complex-valued arrays. This corresponds to a maximum image size of 463×463 with $\kappa=25$ for the traditional FFT upsampling approach, which took 235 s, as compared to 0.78 s with the single-step DFT algorithm. Notice that attempting registration of 2048×2048 images with $\kappa=25$ with the FFT upsampling approach would require over 78 Gbytes of RAM.

The amount of upsampling at which a decrease in computational time was observed for the two-step over the single-step DFT approach was found to be dependent on image size. The intersection occurs approximately at $\kappa=20$, 30, and 45 for image sizes of 64×64, 256×256, and 512×512, respectively. Although the nonlinear optimization approach proved to be the most time consuming in every case, it yields the highest accuracy, as it is not limited by an upsampling factor and can retrieve arbitrary fractional translation values.

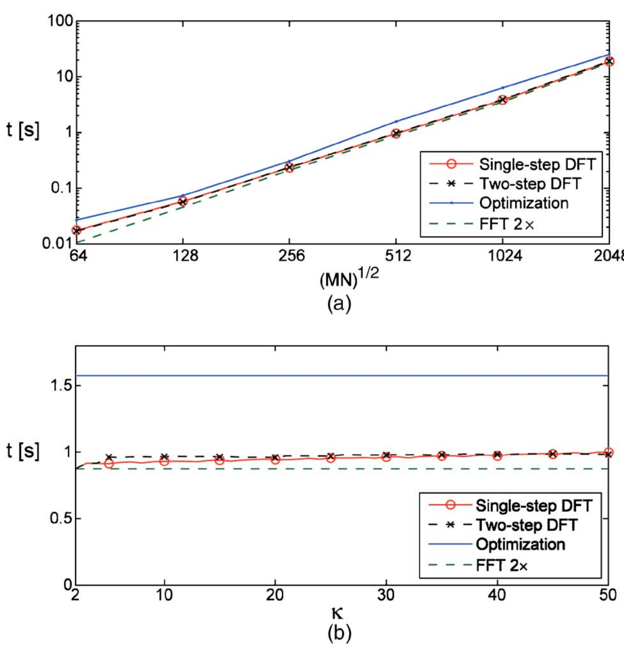

Fig. 2. (Color online) Computation time with respect to (a) image size ($\kappa=25$ for DFT algorithms) and (b) upsampling factor for 512×512 images.

Subpixel image registration by cross correlation is particularly well-suited for comparison of images reconstructed by phase retrieval, where a translated version of the solution that is multiplied by a global complex factor is considered a successful reconstruction. Use of the nonlinear optimization or high upsampling factors is especially important for accurate computation of E in low-noise situations. The accuracy of the translation estimation greatly benefits from large upsampling factors, even in the presence of a moderate amount of noise. The three algorithms can be straightforwardly extended to include variations of the cross-correlation method, such as phase correlation, and to include window functions to account for edge effects [13].

Although other methods (e.g., curve fitting the cross-correlation peak [4] or stochastic sampling approach [8]) can achieve subpixel image registration significantly faster, they sacrifice accuracy in the registration. Our algorithms, as in the FFT upsampling approach, always use all the information available in the images to compute the initial estimate and each point in the upsampled cross-correlation grid, thus rendering them very robust to noise.

The three new registration refinement algorithms were shown to achieve subpixel image registration with the same accuracy as traditional FFT upsampling or better (for the nonlinear optimization algorithm) but with greatly reduced computational time and memory requirements. Use of these algorithms makes accurate registration of large images, even to within a hundredth of a pixel if required, computationally manageable on a regular desktop computer and will prove to be a substantial advantage for any application that requires subpixel image registration.

Portions of this work were presented in [14].

References

1. J. R. Fienup, Opt. Lett. **3**, 27 (1978).
2. J. R. Fienup, Appl. Opt. **21**, 2758 (1982).
3. M. Irani and S. Peleg, CVGIP: Graph. Models Image Process. **53**, 231 (1991).
4. A. R. Wade and F. W. Fitzke, Opt. Express **3**, 190 (1998).
5. M. D. Pritt, J. Opt. Soc. Am. A **10**, 2187 (1993).
6. L. G. Brown, ACM Comput. Surv. **24**, 325 (1992).
7. B. Zitová and J. Flusser, Image Vis. Comput. **21**, 977 (2003).
8. P. Viola and W. M. Wells III, Int. J. Comput. Vis. **24**, 137 (1997).
9. J. R. Fienup, Appl. Opt. **36**, 8352 (1997).
10. W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in Fortran 77* (Cambridge U. Press, 1992).
11. H. H. Barrett and K. J. Myers, *Foundations of Image Science* (Wiley, 2004), p. 172.
12. R. Soummer, L. Pueyo, A. Sivaramakrishnan, and R. J. Vanderbei, Opt. Express **15**, 15935 (2007).
13. S. T. Thurman and J. R. Fienup, Proc. SPIE **6233**, 62330S (2006).
14. M. Guizar-Sicairos, S. T. Thurman, and J. R. Fienup, in *Signal Recovery and Synthesis*, 2007 OSA Technical Digest Series (Optical Society of America, 2007), SMC3.