Signal, Image and Video Processing (2022) 16:1973–1979
https://doi.org/10.1007/s11760-022-02158-7

ORIGINAL PAPER

Subpixel image registration algorithm based on pyramid phase correlation and upsampling

Tianci Li[1,2] · Jianli Wang[1] · Kainan Yao[1]

Received: 10 July 2020 / Revised: 16 January 2022 / Accepted: 21 January 2022 / Published online: 14 February 2022

Abstract

A fast subpixel image registration method is proposed in this paper. The implementation of this method is divided into two steps: coarse registration and fine registration. In the coarse registration stage, we propose a strategy to combine image pyramid with phase correlation; in the fine registration stage, we propose a strategy to perform local upsampling in the frequency domain through matrix multiplication. We compared our algorithm with traditional-feature-based and direct methods, as well as unsupervised learning algorithms. Our empirical results show that compared with traditional methods, our method achieves faster speed, while maintaining equivalent or better accuracy and robustness. In addition, compared with unsupervised learning algorithms, our method can be applied to real-time systems with higher speed requirements, better performance for cases with less overlapping regions, and better robustness to noise.

Keywords Subpixel registration · Pyramid phase correlation algorithm · Image upsampling · Matrix-multiply

1 Introduction

Image registration is the process of overlaying two or more images of the same scene captured at different times from different viewpoints and/or by different sensors [1–4]. This is a crucial process in many image processing applications involving the analysis and integration of multiple images; these applications include scene change detection, image mosaic analysis, and image fusion. Although pixel-level registration may be adequate in many applications, some important problems in remote sensing [5–7], medical imaging [8–12], and biomedical imaging [13–15] have introduced the requirement of high-precision registration. For panoramic target detection, fast, accurate, and reliable image registration is also required.

Direct method and feature-based method are two traditional methods of image registration [16]. Direct methods, such as the Lucas–Kanade algorithm [17], perform image matching by moving or distorting the image and comparing pixel intensity values using error metrics. The robustness of the direct method can be improved by using different performance criteria, such as enhanced correlation coefficient (ECC) [18], representing the image in the Fourier domain [19], and combining the feature-based method with the direct method [20]. The second method is a feature-based method. These methods first use scale-invariant feature transform (SIFT) [21] or other local invariant features to extract feature points in the image. Then, feature matching is used to establish the relationship between the two sets of key points, and RANSAC [22] is used to find the best matching point. These methods have better performance than direct methods, but if enough key points are not detected, or incorrect key point correspondence is generated due to viewpoint difference and illumination between images, this method will fail [23]. In addition, although the feature-based method is much faster than the direct method, the calculation speed of features is still very slow. Although more efficient algorithms such as fast orientation and short rotation (ORB) [24] are proposed, their performance is poor. In recent years, inspired by the success of data-driven depth convolution neural network (CNN) in computer vision, a large number of studies have used CNN methods to estimate optical flow [25–27], dense matching [27,28], depth estimation [29], and homog-

✉ Tianci Li
litianci15@mails.ucas.ac.cn

Kainan Yao
yaokainan001@126.com

1 Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun 130033, China

2 University of Chinese Academy of Sciences, Beijing 100049, China

raphy estimation [30], so as to realize image registration. The most representative unsupervised method [31] improves the traditional supervised learning method by minimizing the pixel-level intensity error measurement that does not need ground real data [30].

Phase correlation is based on the translation property of FT; it transforms the displacement of two correlated images in the spatial domain into a phase difference in the frequency domain. Because only phase information is used, the dependence on image intensity and content is reduced, which makes the method invariant to global linear variations in contrast and brightness. Therefore, phase correlation is robust to frequency-dependent noise, large intensity differences, and time-varying illumination disturbances [32]. Inspired by this, this paper proposes a subpixel image registration algorithm based on pyramid phase correlation and local upsampling. The image pyramid and phase correlation algorithms are combined to achieve coarse registration, and the calculation speed is accelerated by reducing the registration image resolution. For the precise registration stage, this paper proposes a method for local upsampling in the frequency domain through matrix multiplication. Compared with the method of spatial interpolation through zero filling in the frequency domain, this algorithm considerably improves the computational efficiency. The algorithm considers the position of the coarse registration step as a parameter to construct the upsampling matrix and obtains the subpixel translation through the inverse Fourier transform in the matrix form. The algorithm is implemented in the frequency domain, and the inverse fast Fourier transform is applied in the local domain.

We prove that compared with the SIFT algorithm and ECC algorithm, the pyramid phase correlation algorithm achieves faster speed, while maintaining equivalent or better performance for cases with less overlapping regions and better robustness to noise. In addition, our algorithm is faster than unsupervised learning algorithms, so our algorithm can be applied to real-time systems with higher speed requirements and has better robustness.

This paper is organized as follows. In the next section, the theoretical basis of the image registration algorithm is presented. In Sect. 3, the improved subpixel image registration algorithm is described. In Sect. 4, the numerical simulation results are presented. Finally, in Sect. 5, the paper is concluded, and directions for future research are presented.

2 Theoretical basis of image registration algorithm

2.1 Phase correlation algorithm

In processing terms, the correlation between two image signals can be achieved by their convolution, which, in turn, can be used to compare the similarities between the images. In addition, when the image is analyzed in the frequency domain, its Fourier spectrum contains both modulo and phase information. The first represents the gray-level information of the image, whereas the second represents its texture and structure information. For an image $I(x, y)$, its complex spectrum $I_f(u, v)$ includes both amplitude and phase information; in turn, the latter contains the original image position $I(x, y)$. Then, by calculating their correlation, the difference between the images can be obtained, thereby enabling their registration. Therefore, using the phase spectrum matching method, the proposed phase correlation algorithm can solve the image registration problem through simple shifting. The phase correlation algorithm is a popular Fourier domain method used to register two images. It computes a phase difference map that (ideally) contains a single peak. The peak location is proportional to the relative translation $[dx, dy]$ between the two images [33]. This algorithm is equivalent to correlation in the spatial domain, but the calculation is orders-of-magnitude faster in the Fourier domain. The mathematical details are as follows:

Consider two identical images $i1$ and $i2$, with $i2$ shifted by an amount $[\Delta x, \Delta y]$ relative to $i1$:

$$i2(x, y) = i1(x - \Delta x, y - \Delta y). \tag{1}$$

These images satisfy the following periodic boundary conditions:

$$i1(M + x, N + y) = i1(x, y) \tag{2}$$

where the image size is $M \times N$ pixels. Denote the Fourier transforms of $i1$ and $i2$ as $I1$ and $I2$, respectively. From the Fourier shift theorem, $I1$ and $I2$ differ only by a linear phase term $j2\pi(\frac{u\Delta x}{M} + \frac{v\Delta y}{N})$. In particular,

$$I2(u, v) = I1(u, v)e^{-j2\pi\left(\frac{u\Delta x}{M}+\frac{v\Delta y}{N}\right)} \tag{3}$$

where $I2(u, v)$ and $I1(u, v)$ are the corresponding Fourier transforms of $i2(x, y)$ and $i1(x, y)$. The normalized cross-power spectrum of the images, $C12$, is defined as follows:

$$C12(\Delta x, \Delta y) = \frac{I2^{*}_{.}conj(I1)}{|I2^{*}_{.}conj(I1)|} = e^{-j2\pi\left(\frac{u\Delta x}{M}+\frac{v\Delta y}{N}\right)} \tag{4}$$

The operator * is the Schur product (also known as the Hadamard element-by-element matrix product) and $conj$ is the complex conjugate operator.

Equation 5 is an inverse Fourier transform of $C12$. The result is a two-dimensional Dirac delta function $\delta(x - \Delta x, y - \Delta y)$ with a peak location corresponding to the displacement $[\Delta x, \Delta y]$ between the two images. Finally, the coordinates corresponding to the function peak point can be obtained, which, in turn, can be used to calculate the required registration position.

$$\mathcal{F}^{-1}\left(e^{-j2\pi\left(\frac{u\Delta x}{M}+\frac{v\Delta y}{N}\right)}\right)=\delta(x-\Delta x,y-\Delta y) \tag{5}$$

2.2 Image pyramid principle

An image pyramid consists of a sequence of copies of an original image in which both sample density and resolution are decreased in regular steps. The reduced resolution levels are obtained through an efficient iterative algorithm [34,35]. For example, a pixel in an upper layer of an image pyramid summarizes four pixels in the next layer [36]. If multiple pyramid scaling operations are performed, image processing speeds increase exponentially. A five-tap filter was used to generate the image pyramid shown in Fig. 1. The bottom or level zero of the pyramid, G_0, is the original image. This is low-pass-filtered and subsampled by a factor of two to obtain the next pyramid level, G_1. G_1 is then filtered in the same manner and subsampled to obtain G_2. Further repetitions of the filter/subsample steps generate the remaining levels. In particular, the pyramid levels are iteratively obtained as follows. For $0 < l < N$:

$$G_l(i,j)=\sum_m\sum_n w(m,n)G_{l-1}(2i+m,2j+n) \tag{6}$$

The weighting function $w(m,n)$ is referred to as the 'generating kernel'.

3 Improved subpixel image registration algorithm

The subpixel image registration algorithm flow based on pyramid phase correlation and upsampling is shown in Fig. 2. The main steps of the algorithm are as follows:

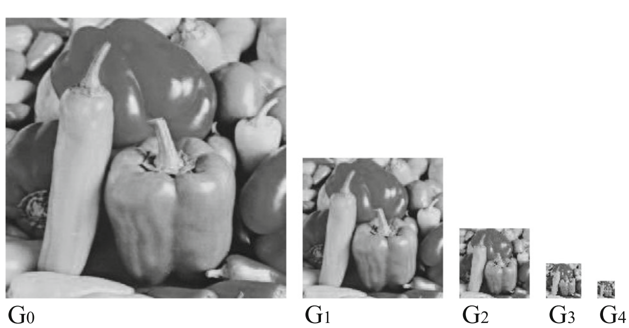

Fig. 1 Gaussian pyramid. Original image, G_0, is repeatedly filtered and subsampled to generate the sequence of reduced resolution images G_1, G_2, and so on. These reduced resolution images are a set of low-pass-filtered copies of the original image in which the bandwidth decreases in one-octave steps

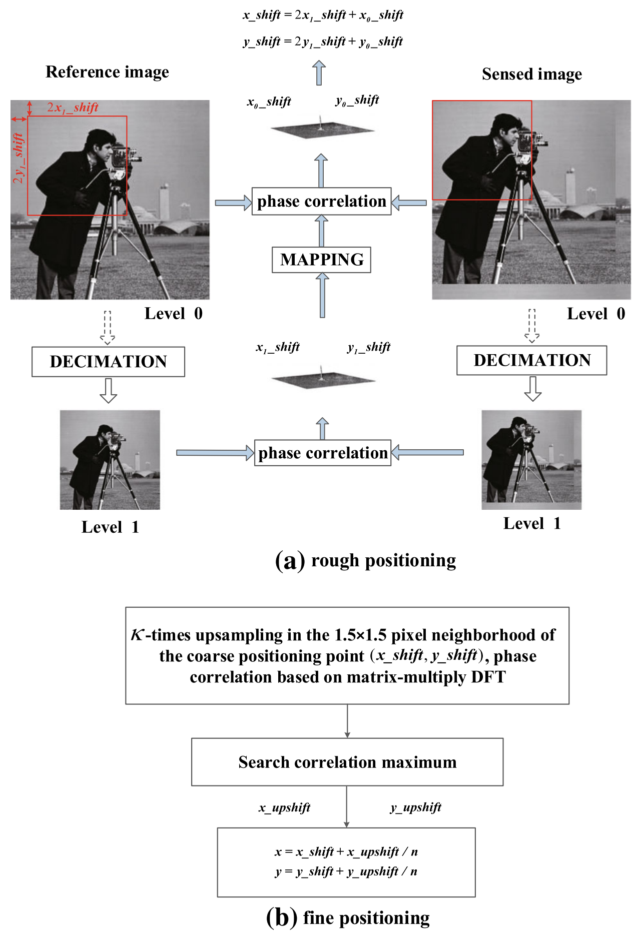

Fig. 2 Flowchart of the subpixel image registration algorithm based on pyramid phase correlation and upsampling

(1) Rough positioning, using the pyramid phase correlation to obtain the translation amount (x_shift, y_shift). Firstly, the reference and sensed images with half the resolution are obtained at level 1 by downsampling these images at level 0, and using the phase correlation to obtain the translation (x_1_shift, y_1_shift). Secondly, phase correlation is used to match the sub-images in the two red boxes in level 0 to obtain the offset (x_0_shift, y_0_shift). In general, the vertices in the top left corner of the red box in the reference image are the vertices offset $(2x_1\_shift, 2y_1\_shift)$, which matches the vertices in the top left corner of the red box in the sensed image. The selection criteria of the red boxes first determine the red boxes position in the sensed image and then determine the red box position of the reference image based on the red box position in the sensed image and the (x_1_shift, y_1_shift). The position of the red boxes in the sensed image can be selected in advance for areas with more textures and obvious gray changes. The selection of the red box is to accelerate the registration speed, and the red frame length is half of the

reference image or sensed image; this selection method can reduce the calculation amount to 1/4. Finally, the pixel-level shift $(x\_shift, y\_shift) = (2x_1\_shift + x_0\_shift, 2y_1\_shift + y_0\_shift)$ is obtained.

(2) Fine positioning. 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 unsampled array, of dimensions $(\kappa M, \kappa N)$ where upsampling factor κ, M, and N are the image dimensions, resulting in an enormous waste of resources if we are interested only in computing an upsampled version of $C12$ 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 $C12$ can be computed within just such a neighborhood without the need to zero-pad $\frac{I2^{*}conj(I1)}{|I2^{*}conj(I1)|}$.

The matrix-multiply DFT uses the properties of the matrix to implement the Fourier transform of the partial point sequence. Given a one-dimensional discrete signal $f(X)$, $X = (x_0, x_1, \ldots, x_{N-1})^{\mathrm{T}}$, Fourier transform of the signal in matrix form is defined as

$$F(U) = E \cdot f(X) \tag{7}$$

where $U = (u_0, u_1, \ldots, u_{N_B-1})^{\mathrm{T}}$ and

$$E = \begin{pmatrix} e^{-j2\pi x_0 u_0} & \ldots & e^{-j2\pi x_k u_0} & \ldots & e^{-j2\pi x_{N-1} u_0} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ e^{-j2\pi x_0 u_k} & \ldots & e^{-j2\pi x_k u_k} & \ldots & e^{-j2\pi x_{N-1} u_k} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ e^{-j2\pi x_0 u_{N_B-1}} & \ldots & e^{-j2\pi x_k u_{N_B-1}} & \ldots & e^{-j2\pi x_{N-1} u_{N_B-1}} \end{pmatrix} \tag{8}$$

Equation 7 is the Fourier transform (FT) in matrix form deduced from the representation by Riemann sum of continuous FT; however, if N_B is equal to N, Eq. 7 is the DFT of the signal.

By extending the one-dimensional matrix-multiply DFT expression directly to the two-dimensional matrix-multiply DFT, the two-dimensional matrix-multiply DFT sequence $f(X, Y)$ can be expressed as a matrix product

$$F(U, V) = E_1 \cdot f(X, Y) \cdot E_2 \tag{9}$$

In the expression: $E_1 = e^{-j2\pi U X^{\mathrm{T}}}$, $E_2 = e^{-j2\pi Y V^{\mathrm{T}}}$; $U = (u_0, u_1, \ldots, u_{N_B-1})^{\mathrm{T}}$, $V = (v_0, v_1, \ldots, v_{N_B-1})^{\mathrm{T}}$; $X = (x_0, x_1, \ldots, x_{N_A-1})^{\mathrm{T}}$, $Y = (y_0, y_1, \ldots, y_{N_A-1})^{\mathrm{T}}$;

According to the matrix Fourier transform in Eq. 9, the inverse Fourier transform of the upsampling matrix is derived, given an input $m = 1.5$ (usually $1 \leq m < 2$), the size of upsampled region, and κ, the precision coefficient in spatial domain, and (x_shift, y_shift), the rough positioning point. The discrete frequency spectrum corresponds to a continuous periodic signal in spatial domain; therefore, in order to obtain a higher-resolution spatial domain signal, we can increase the sampling rate of the signal. It is assumed that the sampling step is $1/\kappa < 1$ pixel. ($\kappa > 0$ is the precision coefficient which can be specified by users.) The initial localization point is in the phase correlation spectrum, and the phase correlation peak may be anywhere in the rectangle with the initial localization point as the center and m as the side length.

The upsampling matrix inverse Fourier transform on L is defined on the basis of the matrix FT:

$$f(X', Y') = \begin{cases} e^{j2\pi X' U^{\mathrm{T}}} \cdot F(U, V) e^{j2\pi V Y'^{\mathrm{T}}} & X' \in L \text{ and } X' \in L \\ 0 & \text{else} \end{cases} \tag{10}$$

Where

$$L = \left\{ (x', y'), \begin{array}{l} x\_shift - \frac{m}{2} \leq x' \leq x\_shift + \frac{m}{2} \\ y\_shift - \frac{m}{2} \leq y' \leq y\_shift + \frac{m}{2} \end{array} \right\} \tag{11}$$

where X' and Y' are spatial distributions of the upsampling, X and Y are spatial distributions of $C12$.

The discrete spectrum corresponds to the periodic signal in the spatial domain. In order to obtain a higher-resolution spatial domain signal, we are here to increase the sampling rate of the signal. The matrix multiplication DFT is used to obtain the κ-time upsampled region in the 1.5×1.5 neighborhood of the rough positioning point (x_shift, y_shift). 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 upsampling pixels). The pixel-level translation is obtained by calculating the phase correlation of the upsampling region $(x\_upshift, y\_upshift)$. Considering the upsampling multiple κ, the translation $(x\_upshift/\kappa, y\_upshift/\kappa)$ of the subpixel is obtained. Therefore, the translation $\Delta = (x, y)$ based on phase correlation and upsampled image registration is

$$\begin{cases} x = x\_shift + x\_upshift/n \\ y = y\_shift + y\_upshift/n \end{cases} \tag{12}$$

In this study, the pyramid phase correlation is used to obtain the rough-positioning point in the original image and the fine-positioning point after κ-time area upsampling. Because the pyramid phase correlation rough positioning has pixel-level precision, the fine-positioning region is set to be 1.5×1.5 centered on the rough-positioning point, to ensure that the precise subpixel-level positioning point is in this region. To obtain higher positioning precision without increasing the calculations, the upsampling multiple κ is taken as 100, and then the inverse Fourier transform is performed in the 150×150 region to obtain the fine-positioning peak, that is, when the positioning precision reaches 0.01 pixels.

4 Numerical simulations

For practical real-time applications, the speed required for an image registration algorithm is often higher for offline processing. In addition, it needs to be strongly robust for cases in which the target is less overlapping regions. In order to verify the performance of our algorithms, it is compared with the traditional-feature-based SIFT algorithm and direct method ECC algorithm, and the newly proposed unsupervised homography estimation algorithm based on deep learning. The ECC direct method and our method are standard Python OpenCV implementation, while the feature-based approaches are Python OpenCV implementations of SIFT RANSAC and the deep learning approaches are implemented in TensorFlow.

The intended use of our algorithm is in panoramic target detection system, which needs high-precision image registration in real time for subsequent target detection and recognition. This requires fast and accurate image registration. We used MS-COCO data to generate test data sets in the same way as in paper [31].

We select 5000 pairs of images from the test data set, of which the image resolution is 320×240, and get the average time-consuming of each registration. The registration time comparison of various algorithms is shown in Fig. 3. Unsupervised CPU and unsupervised GPU are the versions of unsupervised homography estimation algorithm CPU and GPU, respectively. As can be seen from Fig. 3, the time consumption of ECC and SIFT algorithms is much higher than our algorithms and unsupervised learning algorithms. Although the time consumption of unsupervised GPU when using GPU is lower than that of unsupervised CPU, it is still slightly higher than our algorithm. It must be noted that our algorithm only uses the form of two-layer pyramid in the coarse registration stage. For the case of larger image resolution, we can use more layers of pyramid form. Therefore, our algorithm is the most advantageous for real-time registration.

In order to study the effect of overlap rate on matching, we take the 5000 sets of images used earlier as test data and crop each image in two parts. In the subsequent test experiments with different overlap rates, the image on the left is used as a reference image; according to the pixel value ratio of 85% overlap rate, calculate the position of the positioning pixel value and crop out the sensed image of the same size as the reference image in the original image. Similarly, images

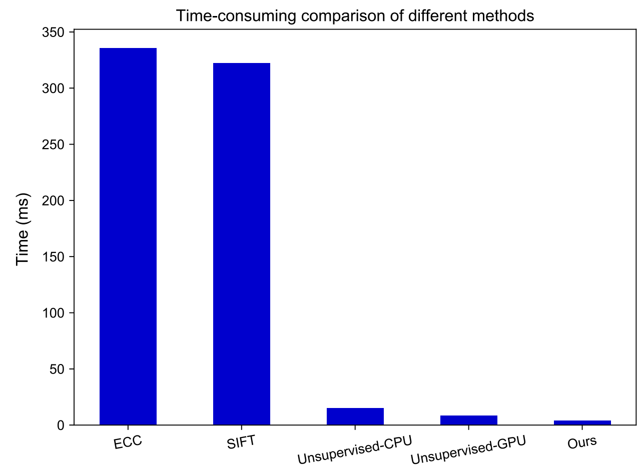

Fig. 3 Time taken by various algorithms to achieve image registration

with 75% and 65% overlap rate can be cropped. The obtained three groups of sensed images are, respectively, tested with the corresponding reference images.

We use the root-mean-square distance (RMSD) between the real result and the calculated result as the error measure. The results of each method are broken down by overlap and performance percentile in Figs. 4, 5, and 6. The unsupervised homography estimation algorithm, SIFT algorithm, and ECC algorithm are used for image registration by estimating homography matrix. Referring to the analysis methods in the paper [31], we break down the results by performance percentiles to illustrate various performance summaries for each method. Specifically, SIFT performs well in 60% of cases, but in the worst 40% of cases, it performs poorly, sometimes completely failing to detect enough features to estimate homography. On the other hand, the unsupervised homography estimation algorithm has more stable performance, but

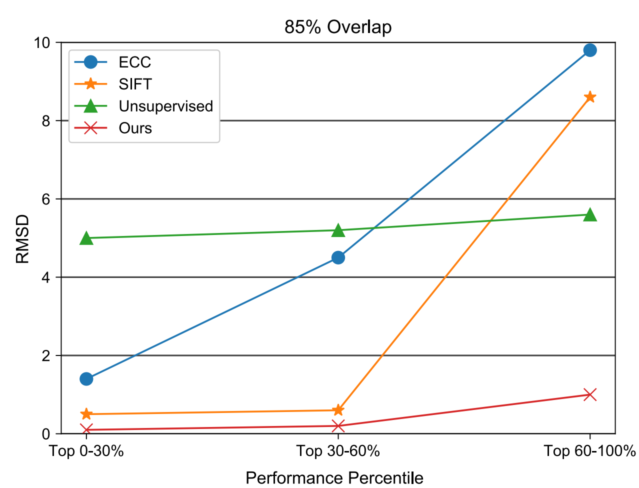

Fig. 4 RMSD values of different performance percentiles for various algorithms at an overlap rate of 85%

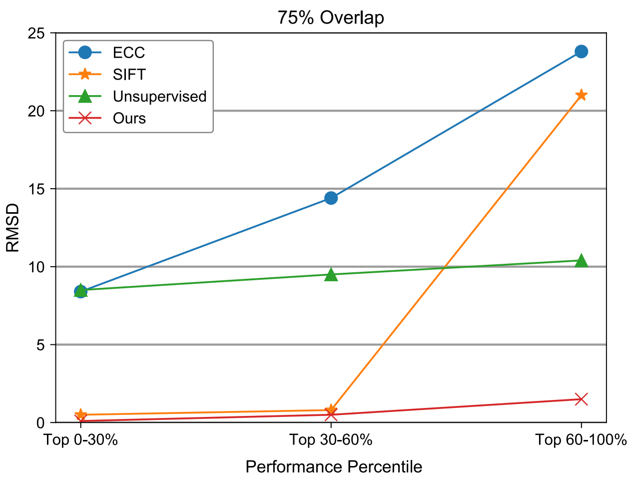

Fig. 5 RMSD values of different performance percentiles for various algorithms at an overlap rate of 75%

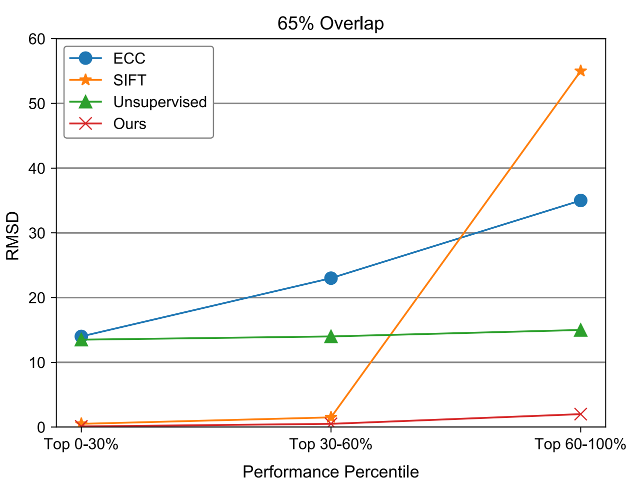

Fig. 6 RMSD values of different performance percentiles for various algorithms at an overlap rate of 65%

the RMSD is still larger. Both the unsupervised homography estimation algorithm and the feature-based method outperform direct methods (ECC), and our method achieves better performance at all performance percentiles. By comparing the RMSD values of the calculation results of each algorithm under different overlap rates, it can be seen that the lower the overlap rate, the worse the performance of various algorithms, but our algorithm can maintain good performance at different overlap rates.

In order to verify the robustness of the algorithm to noise, we add different degrees of Gaussian white noise to the test data. Three groups of data, 5000 pairs of images in each group, were used in the experiment. Their signal-to-noise ratios (SNRs) were 5dB, 20dB, and infinity without noise. The offset of the image is set to subpixel-level offset. The experimental results are shown in Fig. 7. When there is no noise in the image, all algorithms can achieve subpixel registration accuracy. Our algorithm and ECC algorithm have good registration effect even when the SNR is very poor; unsupervised learning algorithm makes the registration result worse with the decrease in SNR; SIFT algorithm is basically unaffected at high SNR, but serious errors will occur at low SNR, because the algorithm mistakenly regards the noise points as feature points. The results show that our algorithm has the best registration performance in the presence of noise.

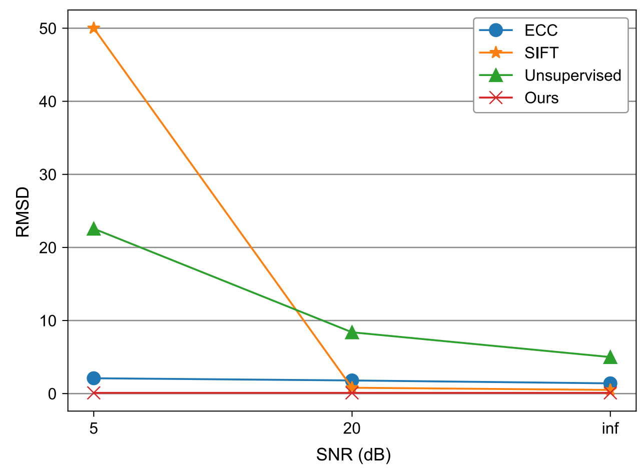

Fig. 7 RMSD values of registration results of various algorithms under different SNRs

5 Conclusions

In this work, a subpixel image registration algorithm is proposed. It uses a pyramid phase correlation algorithm for rough positioning and matrix-multiply upsampling and phase correlation for fine positioning. We compared our algorithm with traditional-feature-based and direct methods, as well as an unsupervised learning algorithms. Our empirical results show that compared with traditional methods, our method achieves faster speed, while maintaining equivalent or better accuracy and robustness. In addition, compared with unsupervised learning algorithms, our method has better performance for cases with less overlapping regions and better robustness to noise. The main advantage of this algorithm is its high speed, which is more suitable for scenes with high speed requirements, such as real-time registration and display of images. Our method is only suitable for image registration situations with only translation at present. For more registration situations, it will be gradually completed in future work. Later, it can be used for remote sensing image registration considering scaling, rotation, projection transformation, and other situations. In the future, it can also be considered to expand to the field of 3D image registration for medical image registration.

References

1. Zitova, B., Flusser, J.: Image registration methods: a survey. Image Vis. Comput. **21**(11), 977–1000 (2003)
2. Bartoli, G.: Image registration techniques: a comprehensive survey. In: Visual Information Processing and Protection Group, pp. 1–54 (2007)
3. Eastman, R.D., Netanyahu, N.S., Moigne, J.L.: Survey of image registration methods. Image Registr. Remote Sens **21**, 35–76 (2011)
4. Deshmukh, M., Bhosle, U.: A survey of image registration. Int. J. Image Process. (IJIP) **5**(3), 245 (2011)
5. Irani, M., Peleg, S.: Improving resolution by image registration. CVGIP Graph. Models Image Process. **53**(3), 231–239 (1991)
6. Shekarforoush, H., Berthod, M., Zerubia, J., Werman, M.: Subpixel Bayesian estimation of albedo and height. Int. J. Comput. Vis. **19**(3), 289–300 (1996)
7. Shekarforoush, H., Chellappa, R.: Data-driven multichannel superresolution with application to video sequences. JOSA A **16**(3), 481–492 (1999)
8. Zhang, Y., Qiao, S., Sun, L., Shi, Q.W., Huang, W., Li, L., Yang, Z.: Photoinduced active terahertz metamaterials with nanostructured vanadium dioxide film deposited by sol–gel method. Opt. Express **22**(9), 11070–11078 (2014)
9. Dong, Y., Long, T., Jiao, W., He, G., Zhang, Z.: A novel image registration method based on phase correlation using low-rank matrix factorization with mixture of Gaussian. IEEE Trans. Geosci. Remote Sens. **56**(1), 446–460 (2017)
10. Wang, C.-W., Ka, S.-M., Chen, A.: Robust image registration of biological microscopic images. Sci. Rep. **4**(1), 1–12 (2014)
11. Lee, C.-Y., Wang, H.-J., Lai, J.-H., Chang, Y.-C., Huang, C.-S.: Automatic marker-free longitudinal infrared image registration by shape context based matching and competitive winner-guided optimal corresponding. Sci. Rep. **7**(1), 1–16 (2017)
12. Guyader, J.-M., Huizinga, W., Poot, D.H.J., van Kranenburg, M., Uitterdijk, A., Niessen, W.J., Klein, S.: Groupwise image registration based on a total correlation dissimilarity measure for quantitative MRI and dynamic imaging data. Sci. Rep. **8**(1), 1–14 (2018)
13. Leprince, S., Barbot, S., Ayoub, F., Avouac, J.-P.: Automatic and precise orthorectification, coregistration, and subpixel correlation of satellite images, application to ground deformation measurements. IEEE Trans. Geosci. Remote Sens. **45**(6), 1529–1558 (2007)
14. Ge, P., Lan, C., Wang, H.: An improvement of image registration based on phase correlation. Optik **125**(22), 6709–6712 (2014)
15. Van den Elsen, P.A., Pol, E.-J.D., Viergever, M.A.: Medical image matching—a review with classification. IEEE Eng. Med. Biol. Mag. **12**(1), 26–39 (1993)
16. Szeliski, R.: Image alignment and stitching: a tutorial. Found. Trends® Comput. Graph. Vis. **2**(1), 1–104 (2006)
17. Lucas, B.D, Kanade, T., et al.: An iterative image registration technique with an application to stereo vision. Vancouver, BC (1981)
18. Evangelidis, G.D., Psarakis, E.Z.: Parametric image alignment using enhanced correlation coefficient maximization. IEEE Trans. Pattern Anal. Mach. Intell. **30**(10), 1858–1865 (2008)
19. Lucey, S., Navarathna, R., Ashraf, A.B., Sridharan, S.: Fourier Lucas–Kanade algorithm. IEEE Trans. Pattern Anal. Mach. Intell. **35**(6), 1383–1396 (2012)
20. Yan, Q., Yi, X., Yang, X., Nguyen, T.: Heask: robust homography estimation based on appearance similarity and keypoint correspondences. Pattern Recogn. **47**(1), 368–387 (2014)
21. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vis. **60**(2), 91–110 (2004)
22. Fischler, M.A., Bolles, R.C.: Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Commun. ACM **24**(6), 381–395 (1981)
23. Wu, F., Fang, X.: An improved RANSAC homography algorithm for feature based image mosaic. In: Proceedings of the 7th WSEAS International Conference on Signal Processing, Computational Geometry and Artificial Vision, vol. 2 (2007)
24. Rublee, E., Rabaud, V., Konolige, K., Bradski, G.: Orb: an efficient alternative to sift or surf. In: 2011 International Conference on Computer Vision, pp. 2564–2571. IEEE (2011)
25. Weinzaepfel, P., Revaud, J., Harchaoui, Z., Schmid, C.: Deepflow: large displacement optical flow with deep matching. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 1385–1392 (2013)
26. Ilg, E., Mayer, N., Saikia, T., Keuper, M., Dosovitskiy, A., Brox, T.: Flownet 2.0: evolution of optical flow estimation with deep networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2462–2470 (2017)
27. Dosovitskiy, A., Fischer, P., Ilg, E., Hausser, P., Hazirbas, C., Golkov, V., Van Der Smagt, P., Cremers, D., Brox, T.: Flownet: learning optical flow with convolutional networks. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2758–2766 (2015)
28. Altwaijry, H., Veit, A., Belongie, S.J., Tech, C.: Learning to detect and match keypoints with deep architectures. In: BMVC (2016)
29. Eigen, D., Puhrsch, C., Fergus, R.: Depth map prediction from a single image using a multi-scale deep network. arXiv preprint arXiv:1406.2283 (2014)
30. DeTone, D., Malisiewicz, T., Rabinovich, A.: Deep image homography estimation. arXiv preprint arXiv:1606.03798 (2016)
31. Nguyen, T., Chen, S.W., Shivakumar, S.S., Taylor, C.J., Kumar, V.: Unsupervised deep homography: a fast and robust homography estimation model. IEEE Robot. Autom. Lett. **3**(3), 2346–2353 (2018)
32. Tong, X., Ye, Z., Yusheng, X., Gao, S., Xie, H., Qian, D., Liu, S., Xiong, X., Liu, S., Luan, K., et al.: Image registration with Fourier-based image correlation: a comprehensive review of developments and applications. IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens. **12**(10), 4062–4081 (2019)
33. Reed, R.A.: Comparison of subpixel phase correlation methods for image registration. Technical report, Arnold Engineering Development Center Arnold AFS TN (2010)
34. Alsaade, F.: Fast and accurate template matching algorithm based on image pyramid and sum of absolute difference similarity measure. Res. J. Inf. Technol. **4**(4), 204–211 (2012)
35. Adelson, E.H., Anderson, C.H., Bergen, J.R., Burt, P.J., Ogden, J.M.: Pyramid methods in image processing. RCA Eng. **29**(6), 33–41 (1984)
36. Burt, P.J., Adelson, E.H. (eds.): The Laplacian pyramid as a compact image code. In: Readings in Computer Vision, pp. 671–679. Elsevier (1987)