2012

Toward Automatic Subpixel Registration of Unmanned Airborne Vehicle Images

Amr Hussein Yousef
Old Dominion University

Jiang Li
Old Dominion University, jli@odu.edu

Mohammad Karim
Old Dominion University

Mark Allen Neifeld (Ed.)

Amit Ashok (Ed.)

Original Publication Citation

Yousef, A. H., Li, J., & Karim, M. (2012) Toward automatic subpixel registration of unmanned airborne vehicle images. In M. A. Neifeld & Amit Ashok (Eds.), *Visual Information Processing XXI, Proceedings of SPIE 8399* (839902). SPIE of Bellingham, WA. https://doi.org/10.1117/12.918935

Toward Automatic Subpixel Registration of Unmanned Airborne Vehicle Images

Amr Hussein Yousef, Jiang Li and Mohammad Karim
Department of Electrical and Computer Engineering
Old Dominion University, Norfolk, VA 23529

ABSTRACT

Many applications require to register images within subpixel accuracy like computer vision especially super-resolution (SR) where the estimated subpixel shifts are very crucial in the reconstruction and restoration of SR images. In our work we have an optical sensor that is mounted on an unmanned airborne vehicle (UAV) and captures a set of images that contain sufficient overlapped area required to reconstruct a SR image. Due to the wind, The UAV may encounter rotational effects such as yaw, pitch and roll which can distort the acquired as well as processed images with shear, tilt or perspective distortions. In this paper we propose a hybrid algorithm to register these UAV images within subpixel accuracy to feed them in a SR reconstruction step. Our algorithm consists of two steps. The first step uses scale invariant feature transform (SIFT) to correct the distorted images. Because the resultant images are not registered to a subpixel precision, the second step registers the images using a fast Fourier transform (FFT) based method that is both efficient and robust to moderate noise and lens optical blur. Our FFT based method reduces the dimensionality of the Fourier matrix of the cross correlation and uses a forward and backward search in order to obtain an accurate estimation of the subpixel shifts. We discuss the relation between the dimensionality reduction factors and the image shifts as well as propose criteria that can be used to optimally select these factors. Finally, we compare the results of our approach to other subpixel techniques in terms of their efficiency and computational speed.

1. INTRODUCTION

Image registration is a process of aligning several images to a reference one or to a common reference coordinate grid. Typically, the alignment process brings the input, or the reference image, into alignment with the base image.[1] In our research we have an optical sensor mounted on a moving platform that captures a sequence of frames with some common area between them. During its flight the UAV experiences roll, pitch and yaw even when it is flying at roughly the same altitude. We define yaw as the rotation of the UAV about an axis pointing directly upwards from the body; roll is a rotation about the axis that connects the UAV tail to its nose; and pitch is a rotation about an axis which is orthogonal to the axes of rotation for pitch and yaw. The change in these parameters changes what falls within the field-of-view (FOV) of the camera[2] which leads to slightly different looks at the same scene. These looks contain similar, but not identical information and should be registered within subpixel accuracy with respect to each others.

Usually, when the UAV experiences yaw, pitch and role the captured image will be affected by shear or it may be tilted or it may have perspective distortion.[3] So, spatial transformations like the affine and the projective can be used to correct for those distortions. The registration step depends on the accurate determination of the the spatial transformation parameters used in the alignment process of the unregistered images. In a previous work,[3] the landmark points that control these spatial transformations are selected manually which is subject to errors and is not powerful. In addition, the corrected UAV images should be registered within subpixel accuracy to minimize the artifacts within the reconstructed SR images and improve its visual quality.

In this paper, we propose a hybrid algorithm to register these UAV images within subpixel accuracy to feed them in a SR reconstruction step. The pipeline of the algorithm consists of two steps. The first step

Contact: Amr Yousef (aabde008@odu.edu) is a graduate student in the Computational Intelligence and Vision Lab, at Old Dominion University (ODU). Jiang Li (JLi@odu.edu) is an assistant professor at ODU. He is affiliated with Virginia Modeling, Analysis, and Simulation Center (VMASC) and is a member of the IEEE and the Sigma Xi. Mohammad Karim (mkarim@odu.edu) is the vice President for Research of ODU.

Visual Information Processing XXI, edited by Mark Allen Neifeld, Amit Ashok, Proc. of SPIE
Vol. 8399, 839902 · © 2012 SPIE · CCC code: 0277-786X/12/$18 · doi: 10.1117/12.918935

automatically selects the landmarks points that control the spatial transformation using SIFT.[4] Then, the second step registers the corrected images using a fast Fourier transform (FFT) based method that is both efficient and robust to moderate noise and lens optical blur and it works efficiently on small shifts between the acquired frames. The improvements depend on reducing the dimensionality of the Fourier matrix of the up-sampled cross correlation. In addition, it uses a forward and a backward search to reduce the required number of complex matrix multiplications that are required to find the peak within the up-sampled cross correlation matrix.

The rest of the paper is organized as follows. In Section 2, An automatic spatial correction of UAV captured images is presented. Frequency domain registration is discussed in Section 3. Section 5 present our simulations and results. Conclusions are summarized in Section 5.

2. AUTOMATIC SPATIAL CORRECTION OF UAV CAPTURED IMAGES

The spatial transformations reallocate the coordinates in one image to new coordinates in another image. If the unregistered image is $f(x, y)$, its registered representation $r(x, y) = T(f(x, y))$ where T represents either the affine or the perspective transformation. The affine mapping is usually utilized when the image is affected by translation, rotation and scaling, i.e., the UAV experiences yaw. If the base image is $b(x, y)$ and the unregistered image is $f(x', y')$ then, given three corresponding points in the two images, the transformation matrix can be calculated from

$$\begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x'_i \\ y'_i \\ 1 \end{bmatrix}, \quad i = 1, \ldots, 3. \tag{1}$$

where (x_i, y_i) and (x'_i, y'_i) are the coordinates of the control points in the base and the unregistered images respectively; a_{11}, a_{12}, a_{21} and a_{22} control the scale, the rotation and the stretch; and a_{13} and a_{23} are the translation in the $x-$ and $y-$ directions.

The projective mapping is usually utilized when the UAV experiences pitch or roll and the acquired images have perspective distortion and are tilted. A mapping h is projective if, and only if, there exists a non-singular 3×3 matrix H such that $h(p) = Hp$, $\forall p$, where $p = [x\ y]$ is the vector representation of a point. If the base image is $b(x, y)$ and the unregistered image is $f(x', y')$ then the projective transformation can be written as

$$x'(h_{31}x + h_{32}y + 1) = h_{11}x + h_{12}y + h_{13} \tag{2}$$

$$y'(h_{31}x + h_{32}y + 1) = h_{21}x + h_{22}y + h_{23} \tag{3}$$

The landmarks points that are used in the extraction of the transformation parameters are selected automatically using the SFIT. The effectiveness of SIFT comes from its robustness to affine changes, noise, illumination change, and 3D view partial change. It consists of four major stages:[4] (1) scale-space creation; (2) keypoint allocation; (3) orientation designation; (4) keypoint descriptor. As shown by Koendernik[5] and Lindeberg[6] the difference-of-Gaussian, (DOG) can be used efficiently in the detection of the stable features inside the image. First, the sift is applied to the reference image and the extracted descriptors will be saved in a database. Then, it's applied to the unregistered images and the extracted descriptors will be matched to the saved database through a nearest neighborhood search that minimizes the Euclidean distance between them.

2.1 Spatial transformations parameters extraction

In our studies, we need to correct the UAV captured images distortions to eliminate the effect of the yaw, pitch, role, and altitude change. Once the SIFT is utilized, the affine transformation will have 6 degree of freedom (DOF) so it needs 3 keypoint matches while the projective transformation has 8 degree of freedom so it needs 4 matches to derive the transformation parameters. With 3 keypoints match, the affine transformation given in Equation (1) can be rewritten as:

$$\begin{bmatrix} x'_1 & y'_1 & 0 & 0 & 1 & 0 \\ 0 & 0 & x'_1 & y'_1 & 0 & 1 \\ x'_2 & y'_2 & 0 & 0 & 1 & 0 \\ 0 & 0 & x'_2 & y'_2 & 0 & 1 \\ x'_3 & y'_3 & 0 & 0 & 1 & 0 \\ 0 & 0 & x'_3 & y'_3 & 0 & 1 \end{bmatrix} = \begin{bmatrix} a_{11} \\ a_{12} \\ a_{21} \\ a_{22} \\ a_{13} \\ a_{23} \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \\ x_2 \\ y_2 \\ x_3 \\ y_3 \end{bmatrix}, \tag{4}$$

which can be written as a system of linear equations $\mathbf{AX} = \mathbf{B}$. The solution of this system can be obtained through the least square method given by the normal equation $\mathbf{X} = [\mathbf{A}^{\mathrm{T}}\mathbf{A}]^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{B}$ which minimizes the sum of the distance between the corresponding locations in the reference and the unregistered images. Similarly, the projective transformation in Equation (3) can be rewritten as

$$\begin{bmatrix} x'_1 & y'_1 & 1 & 0 & 0 & 0 & -x'_1x_1 & -y'_1x_1 \\ 0 & 0 & 0 & x'_1 & y'_1 & 1 & -x'_1y_1 & -y'_1y_1 \\ x'_2 & y'_2 & 1 & 0 & 0 & 0 & -x'_2x_2 & -y'_2x_2 \\ 0 & 0 & 0 & x'_2 & y'_2 & 1 & -x'_2y_2 & -y'_2y_2 \\ x'_3 & y'_3 & 1 & 0 & 0 & 0 & -x'_3x_3 & -y'_3x_3 \\ 0 & 0 & 0 & x'_3 & y'_3 & 1 & -x'_3y_3 & -y'_3y_3 \\ x'_4 & y'_4 & 1 & 0 & 0 & 0 & -x'_4x_4 & -y'_4x_4 \\ 0 & 0 & 0 & x'_4 & y'_4 & 1 & -x'_4y_4 & -y'_4y_4 \end{bmatrix} = \begin{bmatrix} h_{11} \\ h_{12} \\ h_{13} \\ h_{21} \\ h_{22} \\ h_{23} \\ h_{31} \\ h_{32} \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \\ x_2 \\ y_2 \\ \vdots \\ y_4 \end{bmatrix}, \tag{5}$$

which can be solved using the normal equation such as in the affine case.

3. FREQUENCY-DOMAIN BASED REGISTRATION

Up to this point, the UAV images are restored form shear, tilt, and perspective distortions. The following step is to align them within subpixel accuracy which is considered as a crucial step in any SR reconstruction algorithm. Without accurate subpixel registration, severe artifacts will be present within the reconstructed SR images. Many applications require the accuracy of the registration to be within a small portion of a pixel such as in medical imaging, computer vision[7,8] and remote sensing. In the latter, a pixel in Landsat images measures approximately 80 m on the earth, so 0.1 pixel registration accuracy will lead to a resolution of 8 m.[9]

For the case of a translation between two images, the usual technique to address this problem is to compute the cross-correlation between the unregistered and the base images by means of discrete Fourier transform (DFT), and locate its peak.[10] If the image to be registered is $g(x, y)$ and the base image is $f(x, y)$, then the normalized mean square error (NRMSE) E^2 between two images is defined as:[11]

$$\begin{aligned} E^2 &= \min_{\alpha,x_0,y_0} \frac{\sum_{x,y} \left|\alpha g(x - x_0, y - y_0) - f(x, y)\right|^2}{\sum_{x,y} \left|f(x, y)\right|^2} \\ &= 1 - \frac{\max_{x_0,y_0} \left|r_{fg}(x_0, y_0)\right|^2}{\sum_{x,y} \left|f(x, y)\right|^2 \sum_{x,y} \left|g(x, y)\right|^2} \end{aligned} \tag{6}$$

where r_{fg} is the cross-correlation of $f(x, y)$ and $g(x, y)$ defined by:

$$\begin{aligned} r_{fg}(x_0, y_0) &= \sum_{x,y} f(x, y) g^*(x - x_0, y - y_0) \\ &= \sum_{\mu,\nu} \hat{F}(\mu, \nu)\hat{G}^*(\mu, \nu) \exp\left(i2\pi\left(\frac{\mu x_0}{M_1} + \frac{\nu y_0}{M_2}\right)\right). \end{aligned} \tag{7}$$

M_1 and M_2 are the image dimensions; * denotes complex conjugation; and $\hat{F}(\mu, \nu)$ and $\hat{G}(\mu, \nu)$ are the DFTs of $f(x, y)$ and $g(x, y)$ respectively. The 2D-DFT of an image $f(x, y)$ is defined by

$$\hat{F}(\mu, \nu) = \frac{1}{M_1 M_2} \sum_{x=0}^{M-1} \sum_{y=0}^{M_2-1} f(x, y) \exp -i2\pi\left(\frac{x\mu}{M_1} + \frac{y\nu}{M_2}\right) \tag{8}$$

The evaluation of the NRMSE requires solving the more general problem of sub-pixel image registration by locating the peak of cross-correlation $r_{fg}(x, y)$. The usual DFT approach to find the cross-correlation peak to within a fraction, $1/\epsilon$, of a pixel is to

1. compute $\hat{F}(\mu,\nu)$ and $\hat{G}(\mu,\nu)$,
2. embed the product $\hat{F}(\mu,\nu)\hat{G}^*(\mu,\nu)$ in a larger array of zeros of dimension $(\epsilon M, \epsilon N)$,
3. compute the inverse DFT to obtain the up-sampled cross-correlation, and
4. locate its peak.

Although this approach is very accurate and robust to moderate noise but its computational complexity and huge memory requirements make it unrealistic even for small dimension images with large upsampling factors. The computational complexity of this approach is $O\left\{M_1 M_2 \epsilon \left[\log_2(\epsilon M_1) + \epsilon \log_2(\epsilon M_2)\right]\right\}$[10] where ϵ is the upsampling factor and M_1 and M_2 are the image dimensions.

Two algorithms were reviewed and compared against the enhanced speedy proposed approach. The first algorithm is non-linear Optimization gradient routine (NLOGR) and the other one is single step discrete Fourier transform (SSDFT). Both algorithms start with an initial estimate for the location of the cross correlation peak by means of usual FFT approach with upsampling factor ϵ_o of 2 which means that the location of the peak is within ± 0.5 pixel accuracy. Most of the time required for the registration process is consumed in this step which is considered as the main drawback of these algorithms. The SSDFT approach is classified as one of the most reliable and efficient subpixel registration algorithms[12] and it is faster and computationally efficient than NLOGR approach. A speeded-up version of the SSDFT approach is proposed without sacrificing the required accuracy and it is roughly 5X times faster than the original SSDFT approach and can be used efficiently in the case of large dimension images and small subpixel shifts.

3.1 Non-linear Optimization Gradient Routine

The algorithm refines the initial estimate obtained by the usual FFT2X approach using a nonlinear-optimization conjugate-gradient routine to maximize $|r_{fg}(x_0, y_0)|^2$. The partial derivative of r_{fg} with respect to x_0 is given by

$$\begin{aligned} \frac{\partial |r_{fg}(x_0,y_0)|^2}{\partial x_0} &= 2\mathcal{I}\left(r_{fg}(x_0,y_0) \sum_{\mu,\nu} \frac{2\pi\mu}{M_1} \hat{F}^*(\mu,\nu)\hat{G}(\mu,\nu) \right. \\ &\times \left. \exp\left(i2\pi \left(\frac{\mu x_0}{M_1} + \frac{\nu y_0}{M_2} \right) \right) \right) \end{aligned} \tag{9}$$

with a similar expression for the partial derivative with respect to y_0. The 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. Assuming that the usual FFT2X initial estimate is $\mathbf{X}_{(0)} = (x_{(0)}, y_{(0)})$, then the steps to refine this estimate are:

1. compute the gradient of r_{fg} at $(x_{(0)}, y_{(0)})$ as given by

$$\nabla r_{fg}(\mathbf{X}_{(0)}) = \left(\frac{\partial r_{fg}}{\partial x_o}, \frac{\partial r_{fg}}{\partial y_o}\right)|_{(x_{(0)}, y_{(0)})} \tag{10}$$

2. start with $\mathbf{d}_{(0)} = \mathbf{r}_{(0)} = \nabla r_{fg}(\mathbf{X}_{(0)})$.
3. find $\alpha_{(i)}$ that minimizes $\nabla |r_{fg}(\mathbf{X}_{(i)} + \alpha_{(i)}\mathbf{d}_{(i)})|^2$
4. update:
   (a) $\mathbf{X}_{(i+1)} = \mathbf{X}_{(i)} + \alpha_{(i)}\mathbf{d}_{(i)}$,
   (b) $\mathbf{r}_{(i+1)} = \nabla r_{fg}(\mathbf{X}_{(i+1)})$

5. calculate $\beta_{(i+1)}$ as given by

$$\beta_{(i+1)} = \frac{\mathbf{r}_{(i+1)}^T \mathbf{r}_{(i+1)}}{\mathbf{r}_{(i)}^T \mathbf{r}_{(i)}} \tag{11}$$

6. calculate $\mathbf{d}_{(i+1)}$ as given by

$$\mathbf{d}_{(i+1)} = \mathbf{r}_{(i+1)} + \beta_{(i+1)} \mathbf{d}_{(i)} \tag{12}$$

7. stop when the maximum iterations exceeded a certain number or $\left\|\mathbf{r}_{(i)}\right\| \leq \epsilon \left\|\mathbf{r}_{(i+1)}\right\|$ with $\epsilon < 1$.

The parameter $\alpha_{(i)}$ can be obtained using general line search using Newton-Raphson Method. If we let $\nabla |r_{fg}(\mathbf{X}_{(i)} + \alpha_{(i)}\mathbf{d}_{(i)})|^2 = f(x + \alpha d)$, then the Taylor expansion of the function f is given by

$$\begin{aligned} f(x+\alpha d) &\approx f(x) + \alpha \left[\frac{d}{d\alpha} f(x+\alpha d)\right]_{\alpha=0} + \frac{\alpha^2}{2}\left[\frac{d^2}{d\alpha^2} f(x+\alpha d)\right]_{\alpha=0} \\ &= f(x) + \alpha \left[f'(x)\right]^T d + \frac{\alpha^2}{2} d^T f''(x) d \end{aligned} \tag{13}$$

Differentiating this equation with respect to α yields

$$\frac{d}{d\alpha} f(x+\alpha d) \approx \left[f'(x)\right]^T d + \alpha d^T f''(x) d \tag{14}$$

The function $f(x + \alpha d)$ is minimized by setting $\frac{d}{d\alpha} f(x + \alpha d)$ to zero and hence

$$\alpha = -\frac{f'^T d}{d^T f'' d} \tag{15}$$

where $f''(x)$ is the Hessian matrix defined by

$$f''(x) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1} & \frac{\partial^2 f}{\partial x_1 \partial x_2} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2 \partial x_2} \end{pmatrix} \tag{16}$$

3.2 Single step discrete Fourier transform approach

The second efficient subpixel algorithm was developed by Guizar-Sicairos et al.[10] to register images with the same accuracy obtained by the usual FFT approach but with a huge reduction in computational time and memory requirements. Their technique was classified as one of the most reliable and best algorithms to register images using phase correlation methods.[12] The SSDFT works on two steps. The first step, which is similar to NLOGR, finds an initial estimate for the location of the cross correlation peak between two images using the usual FFT approach with upsampling factor of $\epsilon_0 = 2$. The usual FFT approach causes a tremendous waste of memory and processing time as it must process the entire zero padded upsampled matrix of dimensions $(\epsilon M_1, \epsilon M_2)$ to get the accurate peak location. On the contrary, the SSDFT approach searches for the accurate peak in a small window around the initial estimate by means of DFT instead of FFT. It utilizes the DFT implementation to obtain an upsampled version of the cross correlation in a small window of size $1.5\epsilon \times 1.5\epsilon$ around the initial estimate without zero padding the product $\hat{F}(\nu, \omega)\hat{G}^*(\nu, \omega)$. This process is implemented by rewriting Equation (8) as a product of three matrices of dimensions $(1.5\epsilon, M_1)$, (M_1, M_2), and $(M_2, 1.5\epsilon)$. Then a search for the peak is done over the output matrix of size $(1.5\epsilon, 1.5\epsilon)$. The computational complexity of this approach is $O(M_1 M_2 \epsilon)$ which is a great improvement to the usual FFT approach.

Although the SSDFT approach is an efficient subpixel registration algorithm, its main disadvantage is that most of the time needed for registration is consumed in searching for the initial estimate. This drawback will be tremendously improved in our proposed approach which greatly reduces the time required for locating the initial estimate and also reduces the time required for the refinement step.

Table 1: Optimum sampling factors for different image sizes.

Image size	Sampling factor K
128	4
256	8
512	16
1024	32
2048	64
4096	128
8192	256

3.3 Speeded-up SSDFT approach

We enhanced the SSDFT approach by reducing the required time for its two steps, i.e., the time required for the initial estimation of the peak location; and the time needed for the refinement step. Our approach to reduce the computational time for estimating the initial peak location depends on reducing the dimension of the Fourier transform of the cross correlation matrix and by applying the inverse FFT, the initial estimate can be obtained in a faster way than the SSDFT approach.

Suppose we replaced the Fourier transform of both the reference image $\hat{F}_1(\nu,\omega)$ and the unregistered version $\hat{F}_2(\nu,\omega)$ by a sampled version of them then the right hand side of Equation (8) becomes

$$\sum_{\mu,\nu}\sum_{m,n} \hat{F}(\mu,\nu)\hat{G}^*(\mu,\nu)\delta(\mu - K_1 m)\delta(\nu - K_2 n) \times \exp\left[i2\pi\left(\frac{d_x\mu}{M_1} + \frac{d_y\nu}{M_2}\right)\right]; \quad (17)$$

where K_1 and K_2 are the sampling factors along the $x-$ and $y-$ directions respectively and δ is the Dirac delta function. Rearranging the sums and by using the sifting property of the Dirac delta function, the last equation can be written as

$$\sum_{m,n} \hat{F}(K_1 m, K_2 n)\hat{G}^*(K_1 m, K_2 n) \times \exp\left[i2\pi\left(\frac{d_x m}{M_1/K_1} + \frac{d_y n}{K_2/M_2}\right)\right]; \quad (18)$$

which represents the inverse Fourier transform of the product at the new reduced dimension. So, the idea here is to sample the Fourier transform of the two images being registered and to apply the same SSDFT approach in searching for the initial estimate of the peak location. The computational complexity of our enhancement is $O(K_1^{-1}K_2^{-1}M_1M_2\epsilon)$ which is a great improvement over the SSDFT approach.

The selection of the sampling factors K_1 and K_2 will depend on the image size which are preferred to be of base of 2 to gain the full power of the FFT. Table 1 shows the sampling factors for different images sizes. If K_{128} is the sampling factor for an image of size 128×128 then the sampling factor for an image of size $M \times M$ can be extracted as $K_M = M/128 * K_{128}$. These sampling factors can make the computational time required for the initial estimate of the peak location to be roughly the same. To avoid aliasing, K_i with $i = \{1 \text{ or } 2\}$ must be selected such that it's less than $M/(2\tilde{d}_i)$ where $\tilde{d}_i$ is the expected subpixel shifts in the $x-$ or $y-$ directions that depends on the application.

The second enhancement of the SSDFT approach reduces the number of matrix multiplications required to find the accurate location in the upsampled cross correlation window by minimizing the number of matrix multiplications required to obtain a partial inverse DFT matrix. Consider Equation (8) to be written as a product of three matrices as given by

$$C_{f_1 f_2} = A_{1.5\epsilon \times M_1} * B_{M_1 \times M_2} * C_{M_2 \times 1.5\epsilon}, \quad (19)$$

then the SSDFT approach searches for the accurate location by multiplying the whole three matrices and then search for the peak in the resultant matrix $C_{f_1 f_2}$ which consumes sometime as not all the values inside the output matrix are required.

In our approach we overcome this weakness by partially obtaining some rows or columns inside the resultant matrix and accurately determine the peak location using a forward and backward search. Even the peak of the up-sampled cross correlation is close to one of the borders of the resultant matrix or not, the proposed algorithm reduces the number of complex matrix multiplications that are required to find the accurate estimation of the peak location. Figure 1 shows an upsampled cross correlation version of a window of size 150×150 (upsampling factor =100) around the initial peak location and Figure 2 shows its level curves. By calculating partial parts

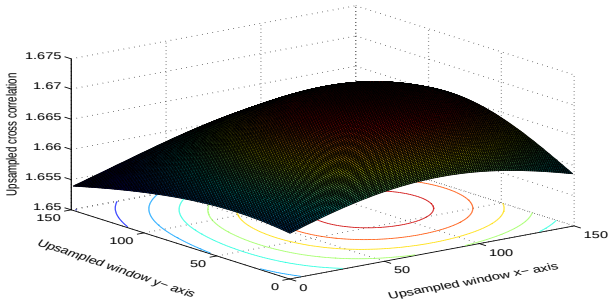


Figure 1: Upsampled cross correlation with window size 150×150.

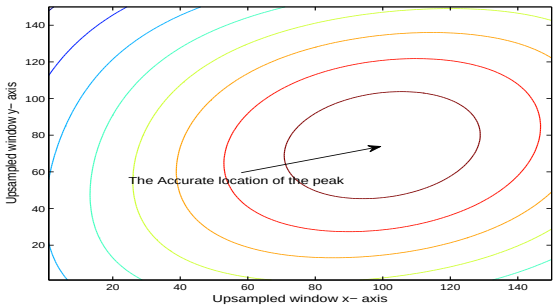


Figure 2: Level curves for the upsampled cross correlation.

of upsampled window, we can speed up the search for the accurate peak location. This can be done using the following steps:

1. calculate the output matrix borders R_1, R_{M_1}, C_1, and C_{M_2} as defined by

$$\begin{aligned} R_i &= (A(i,:) * B) * C, \quad i = 1 \text{ or } M_1 \\ C_j &= A * (B * C(:,j)), \quad j = 1 \text{ or } M_2 \end{aligned} \tag{20}$$

   where the use of "()″ specify the order of matrix multiplication,

2. find the max value across these borders and assume for example it is across the first row $R(1,:)$,

3. from $R(1,:)$ the algorithm starts a forward search in steps of σ to find the next max value across the rows $R(\sigma i + 1,:)$ where i refers to the iteration number and it stops when the next maximum value drops below the previous one.

4. from the last scanned row, the algorithm starts a backward search with a decrement of 1 until the next maximum value is less than the previous one.

At the last scanned row, the algorithm accurately finds the maximum peak and its location.

Table 2: Total time required to register images of dimension 512×512 with different subpixel accuracy.

Subpixel accuracy (pixels)	Enhanced SSDFT (seconds)	SSDFT (seconds)
0.1	0.15	0.54
0.01	0.2	0.63
0.001	1.4	2.98
0.0001	75.13	151.56

4. SIMULATIONS AND RESULTS

In our simulations, we used 3 sets of images that are used to simulate the UAV captured images. These images contain the same scene but with different views. Figure 3 shows the results of applying the SIFT to a set of home images with different orientations to simulate the distortions of shear, tilt and rotations. Also, Figure 4 and 5 show the results of registering two different sets of aerial images.[13] The left column in these three figures show the corresponding keypoints or landmarks matches between the reference images and the unregistered ones while the right column show the images after the registration with dark regions showing the areas of overlapping between these processed images that can be used to reconstruct SR images. Once the appropriate transformation is applied to the unregistered image, the reprojected images will be registered within subpixel accuracy using the enhanced SSDFT approach.

To evaluate our improvements against the SSDFT approach in terms of the speed of computations, we compare the performance of the two approaches against different image sizes ranging from 256×256 up to 8192×8192 in a multiple of 2. The images are corrupted by additive white Gaussian noise and blurred by a Gaussian kernel to simulate the optical lens blur. Also, they are shifted by a lateral shifts of (3.48574,7.73837) in pixels to obtain the unregistered versions of them. The simulations are performed using MATLAB 7.8 Release 2009a program on OPTIPLEX 780 (Intel(R) Core (TM)2 Quad 2.66 GHz CPU, 8.00 GB RAM, MS Windows 7 Professional 2009). The SSDFT technique and the enhanced one register images with accuracy of 0.01 pixel. The estimated shifts for both algorithms are (3.49, 7.74) with estimation error $\delta s = 0.00456$ of a pixel where the estimation error is given by the l_2-norm between the actual shifts (d_x, d_y) and the estimated shifts $(\tilde{d}_x, \tilde{d}_y)$ as given by

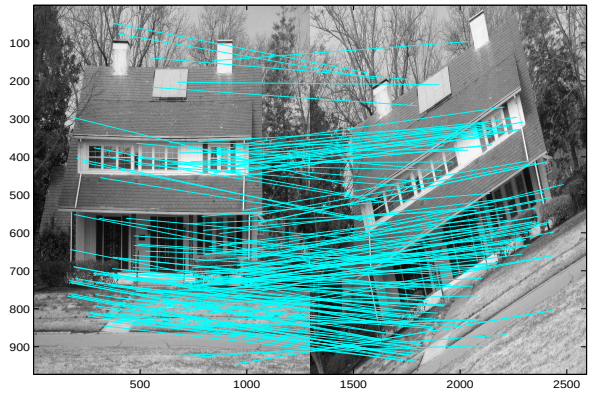
$$\delta s = \sqrt{(d_x - \tilde{d}_x)^2 + (d_y - \tilde{d}_y)^2} \tag{21}$$

Figure 6 and 7 show the comparison between the SSDFT approaches (the original and the enhanced one) and the NLOGR in terms of the estimated NRMSE $\hat{E}$ and the estimated error δs. The greater the upsampling factor the better is the estimate for the subpixel shifts and the lower value of the NRMSE. A comparison of the computational time required for the initial peak estimation, the refinement step, and the total registration time are shown in Figure 8, Figure 9, and Figure 10 respectively. It can be seen that our approach greatly reduces the amount of time required to obtain the initial estimate for the peak location compared to the SSDFT approach. Also, our approach makes the time required for this step roughly the same regardless of the image size which can be done by controlling the selection of the sampling factors K_1 and K_2. For example, for 8192×8192 image it requires around 2 milliseconds using our approach while it needs around 200 seconds using the SSDFT approach. In the refinement step, our method enhances the performance of the SSDFT approach as can be seen in Figure 9. Through out simulations, we set the forward step σ to 0.3ϵ. The total computational time required for the whole registration process for both approaches is shown in Figure 10. Over all, our approach is approximately 5X faster than the SSDFT approach.

Also, we test the performance of both algorithms against changing the required subpixel accuracy which can be seen in Table 2. Our enhancement increases the attainable subpixel accuracy with a large decrease in the computational time.

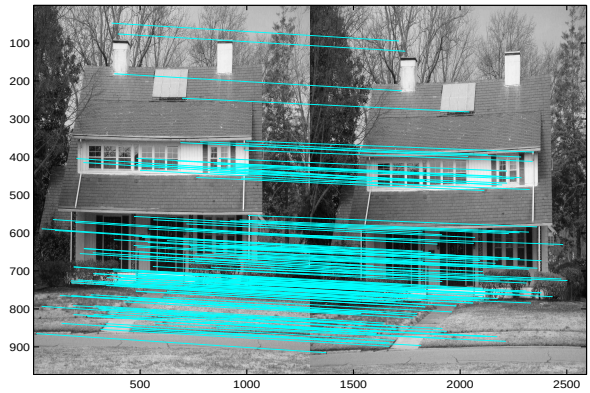
5. CONCLUSIONS

In this paper we present a hybrid subpixel registration algorithm that can be applied efficiently on the UAV captured images and it can also be used to register a set of images with a common overlap between them and have



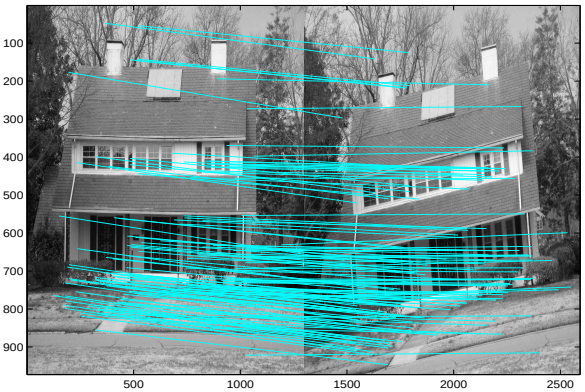
(a) Matched features between base and LR1

(b) Corrected LR1



(c) Matched features between base and LR2

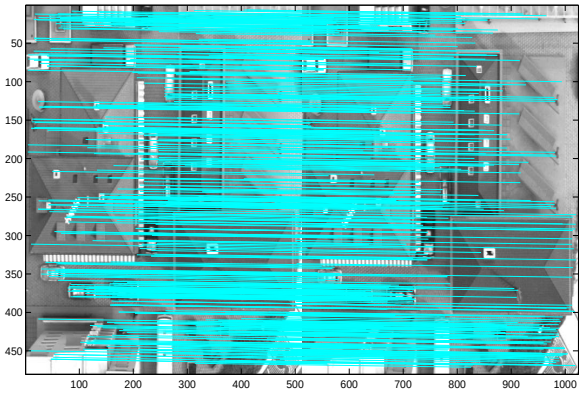
(d) Corrected LR2



(e) Matched features between base and LR3

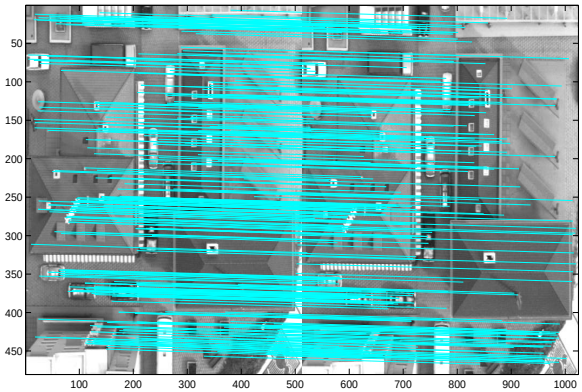
(f) Corrected LR3

Figure 3: Corrected LR images after using the SIFT algorithm (Home images).

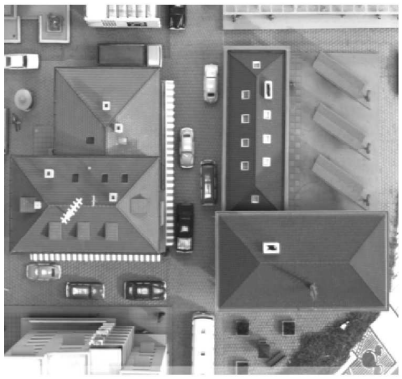


(a) Matched features between base and LR1

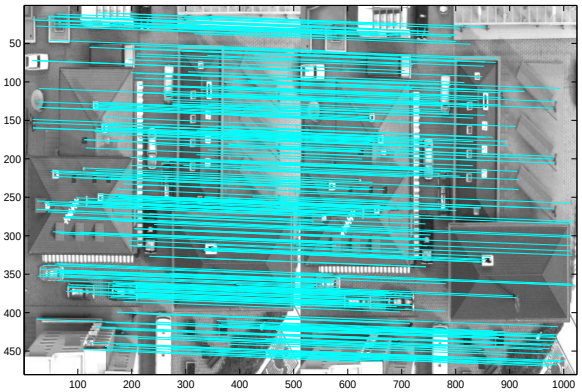
(b) Corrected LR1



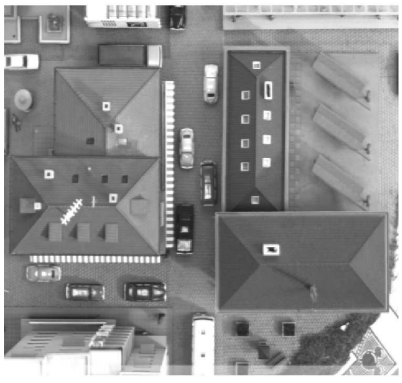
(c) Matched features between base and LR2



(d) Corrected LR2

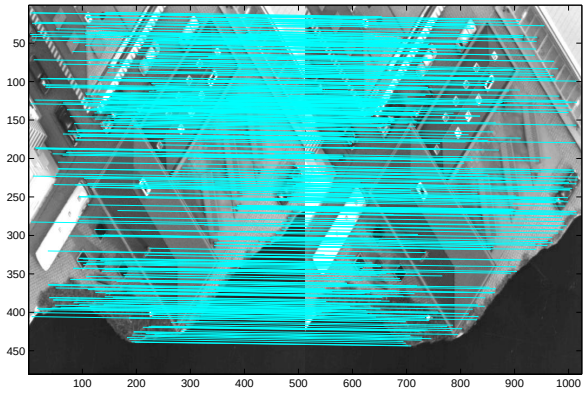


(e) Matched features between base and LR3



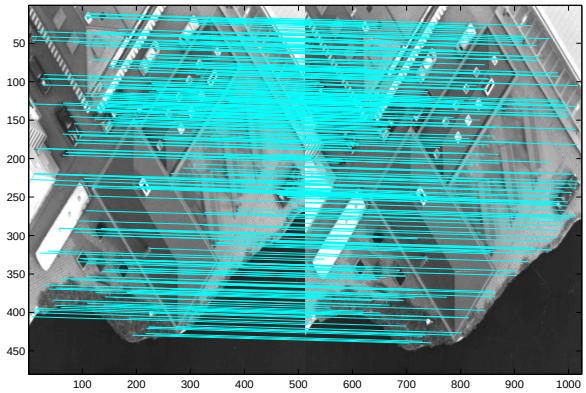
(f) Corrected LR3

Figure 4: Corrected LR images after using the SIFT algorithm (Aerial images 1).

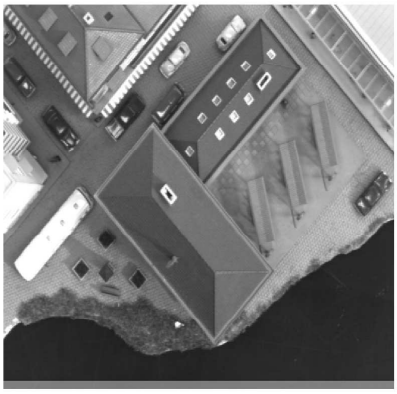


(a) Matched features between base and LR1

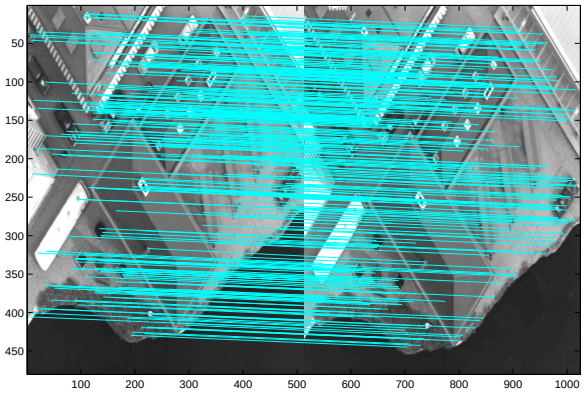
(b) Corrected LR1



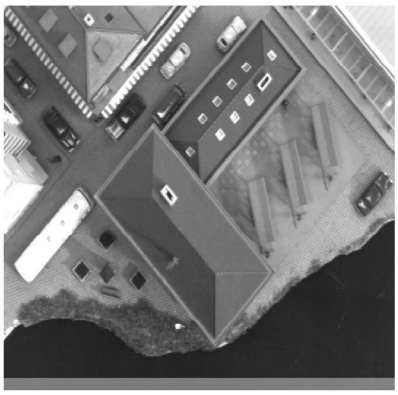
(c) Matched features between base and LR2



(d) Corrected LR2



(e) Matched features between base and LR3



(f) Corrected LR3

Figure 5: Corrected LR images after using the SIFT algorithm (Aerial images 2).

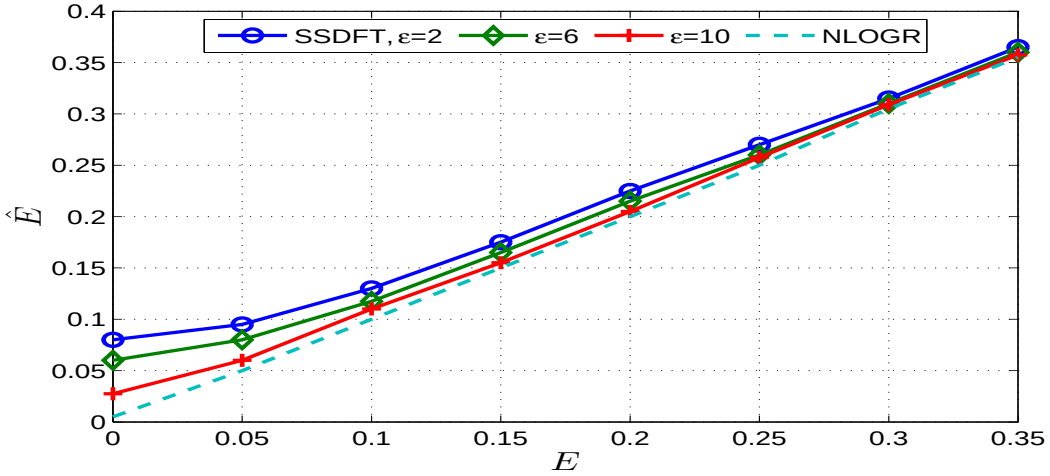


Figure 6: The estimated NRMSE $\hat{E}$ against the actual one.

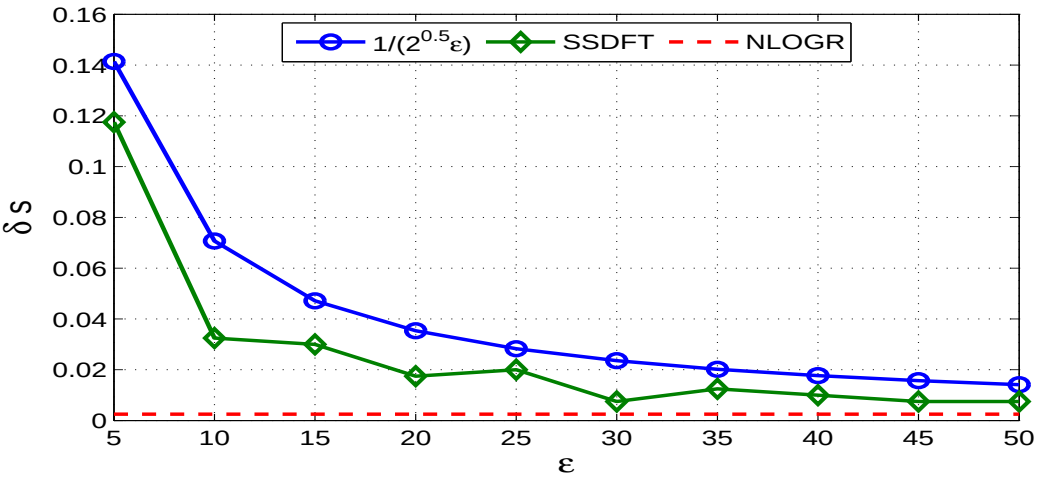


Figure 7: The upsampling factor ϵ against the estimation error δs.

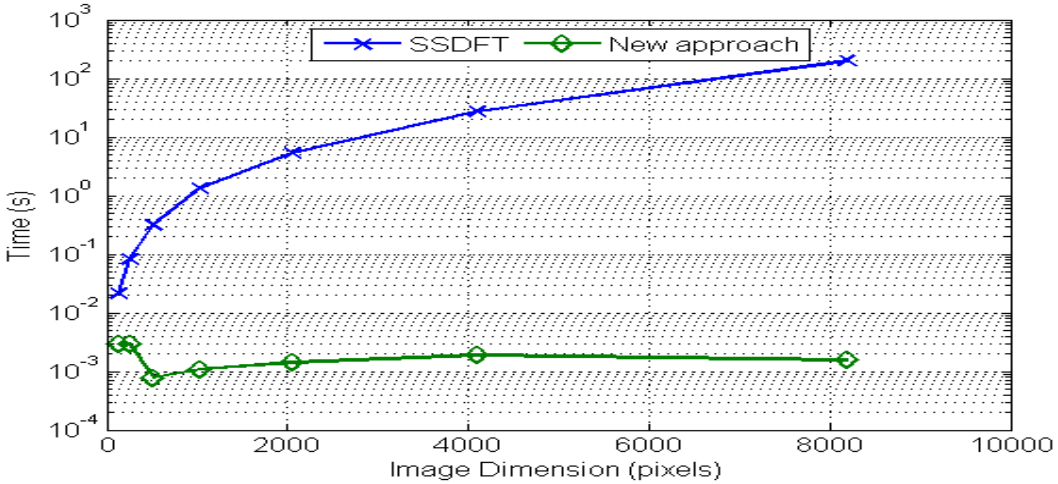


Figure 8: Computational time required for the initial estimate.

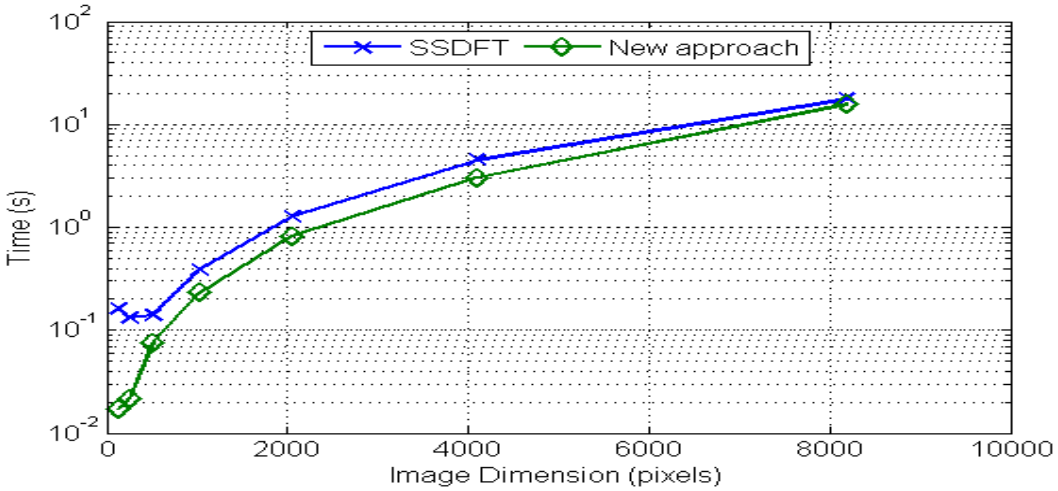


Figure 9: Computational time for the refinement step.

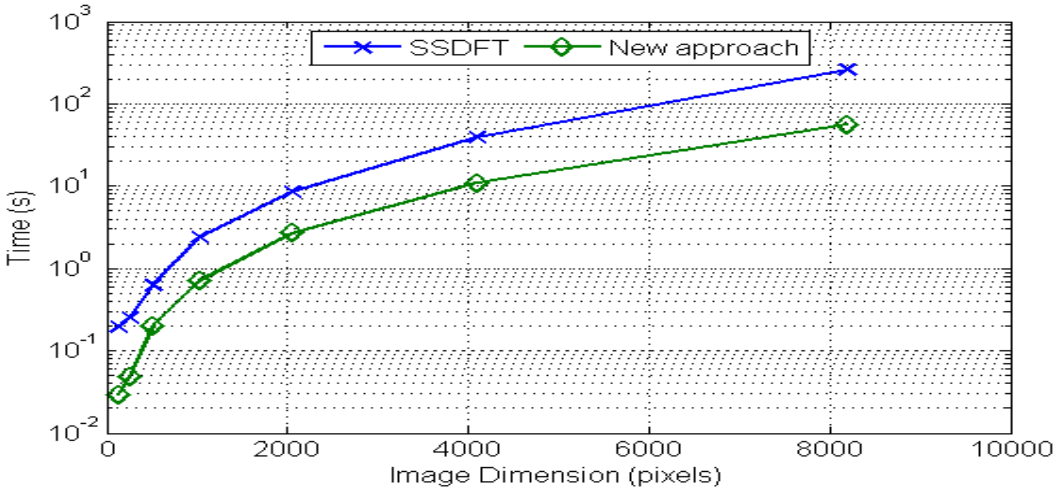


Figure 10: Total time required for registration.

arbitrary motion. The proposed algorithm works on two phases. The first phase correct the acquired images from shear, tilt and scaling that result from the different rotational parameters. The second phase accurately register the corrected frames within subpixel accuracy by reducing the dimensionality of the upsampled cross correlation matrix and reduce the number of complex matrix multiplications that are required for correct estimation of subpixel shifts. The proposed enhanced approach registers images that differ by translational shifts or scaled by a constant and can be used in the presence of moderate noise. Our enhanced approach offers a great reduction in computational time and memory requirements against the SSDFT and the usual FFT approaches without sacrificing the required accuracy. It can register very large dimension images of size 8192×8192 in roughly 1 minute compared to 5 minutes using SSDFT approach with subpixel accuracy of 0.01 pixel. Other subpixel registration techniques such as cross correlation surface fitting[8] or stochastic sampling approaches[14] can perform faster but their registration accuracy is not as efficient.

ACKNOWLEDGMENTS

The authors wish to thank the NASA Aviation Safety Program for the funding which made this work possible. In particular, Dr. Li's work was partially supported under NASA cooperative agreement NNL07AA02A.

REFERENCES

1. M. D. Pritt, "Image registration with use of the epipolar constraint for parallel projections," *J. Opt. Soc. Am. A* **10**, pp. 2187–2192, Oct 1993.
2. Z. ur Rahman, G. D. Hines, and M. J. Logan, "Detecting changes in terrain using unmanned aerial vehicles," *Visual Information Processing XIV* **5817**(1), pp. 53–63, SPIE, 2005.
3. A. H. Yousef and Z. ur Rahman, "Super-resolution reconstruction of images captured from airborne unmanned vehicles," in *Visual Information Processing*, Z. ur Rahman, S. E. Reichenbach, and M. A. Neifeld, eds., *SPIE Proceedings* **7701**, p. 77010, SPIE, 2010.
4. D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vision* **60**, pp. 91–110, November 2004.
5. J. Koenderink, "The structure of images," *Biological Cybernetics* **50**, pp. 363–370, 1984. 10.1007/BF00336961.
6. T. Lindeberg, "Scale-space theory: A basic tool for analysing structures at different scales," *Journal of Applied Statistics* , pp. 224–270, 1994.
7. M. Irani and S. Peleg, "Improving resolution by image registration," *CVGIP: Graph. Models Image Process.* **53**, pp. 231–239, April 1991.
8. A. Wade and F. Fitzke, "A fast, robust pattern recognition asystem for low light level image registration and its application to retinal imaging," *Opt. Express* **3**, pp. 190–197, Aug 1998.
9. Q. Tian and M. N. Huhns, "Algorithms for subpixel registration," *Comput. Vision Graph. Image Process.* **35**, pp. 220–233, August 1986.
10. M. Guizar-Sicairos, S. T. Thurman, and J. R. Fienup, "Efficient subpixel image registration algorithms," *Optics Letters* **33**, pp. 156–158, 2008.
11. J. R. Fienup, "Invariant error metrics for image reconstruction," *Appl. Opt.* **36**, pp. 8352–8357, Nov 1997.
12. R. A. Reed, "Comparison of subpixel phase correlation methods for image registration," final report, Aerospace Testing Alliance, Arnold Air Force Base, Tennessee, USA, April 2010.
13. C.-C. Wang, "Vision and autonomous systems center's image database." *http://vasc.ri.cmu.edu/idb/*.
14. P. Viol and W. M. W. III, "Alignment by maximization of mutual information," *Int. J. Comput. Vis* **24**(2), p. 137154, 1997.