Comparison of stochastic gradient-based optimization techniques for nonlinear satellite image coregistration problem

S. Manthira Moorthi[1,2,]* and R. Sivakumar[3]

[1]Signal and Image Processing Area, Space Applications Centre, Indian Space Research Organisation, Ahmedabad 380 015, India
[2]Department of Physics and Nanotechnology, SRM University, Kattankulathur 603 203, India
[3]Department of Civil Engineering, SRM University, Kattankulathur 603 203, India

Information-oriented intensity-based cost functions are generally used for optimization frameworks in automatic satellite image registration. Optimization mechanics which updates the transform parameters in the iterative loop requires estimation of derivatives of the cost function to set-up update rules that retrieve the deformation model between the image pairs. Application of stochastic approximation of cost function and its derivatives for solving optimization problems while the objective function is non-differentiable or non-smooth or computed with noise is encountered in real-world problems. The known methods of approximation for solving these problems use the idea of stochastic gradient and certain rules of changing the step length for ensuring convergence. In this article, satellite image coregistration problem is chosen for comparing the performance of two important stochastic optimizers like adaptive stochastic gradient descent and simultaneous perturbation stochastic approximation. Coregistration datasets from Resourcesat-2 LISS-4 MX sensor are chosen for different terrains and features to study subpixel accuracies of order better than 1/20th of a pixel achieved in the comparison of two different optimization techniques employed in intensity-based automatic image registration framework.

Keywords: Coregistration problem, remote sensing, satellite image, simultaneous perturbation, stochastic optimization.

INTENSITY-based image registration optimization framework (Figure 1) uses similarity metric such as mutual information (MI) as an objective function. The continuous optimization loop executes to retrieve the geometric transformation parameters to overlay an image pair on each other. It updates the parameters using a stepsize and direction estimated in further using the gradient of the objective function. The objective function measurements could be noisy due to various factors in the imaging context such as data occlusion and unsteady or uncharacterized platforms. In such cases, stochastic modelling of the objective function, its gradient or gradient approximation using the noisy input data can be a valuable tool[1–6]. Algorithms where the gradients are estimated through explicit derivation such as Robbins–Monro approximation are faster and also capable of avoiding local minima[1]. However, there are cases where explicit derivation and estimation of gradients can be difficult or complex. Even otherwise it is important to use a technique for gradient approximation without explicit derivatives. A comparison is required between explicit and non-explicit cases, which is done in this study with more emphasis on the latter case. To state explicitly, in this article adaptive stochastic gradient descent (ASGD) and simultaneous perturbation stochastic approximation (SPSA) optimization schemes are compared for their performance in the overall image registration framework (Figure 1).

Klein et al.[7] compared the performance of many optimization methods, which includes gradient descent approaches with two different stepsize selection algorithms in the medical image registration domain. In the remote sensing image registration domain, evaluation/comparison of optimization methods was performed by Manthira Moorthi et al.[8]. Cole-Rhodes et al.[9] examined the performance of SPSA with MI on satellite image registration. Li et al.[10] have highlighted the performance of SPSA with MI on optical and microwave images obtained from satellite platforms. The present study compares the performance of ASGD and SPSA in the Resourcesat-2 LISS-4 multispectral image coregistration problem recently reported by Manthira Moorthi et al.[11], where ASGD approach was discussed in detail, including estimation of MI, its derivative, optimization mechanics and deformation models handled by the registration approach. The comparison addresses relative performance of the stochastic optimization techniques with respect to objective function or the similarity measure search space, transformation choices, gradient estimation, optimization mechanics and the initial transform setting for a variety of landforms encountered in satellite images. Different data regions such as urban, land and sea, undulating terrains, and occluded images are used for exercises to study
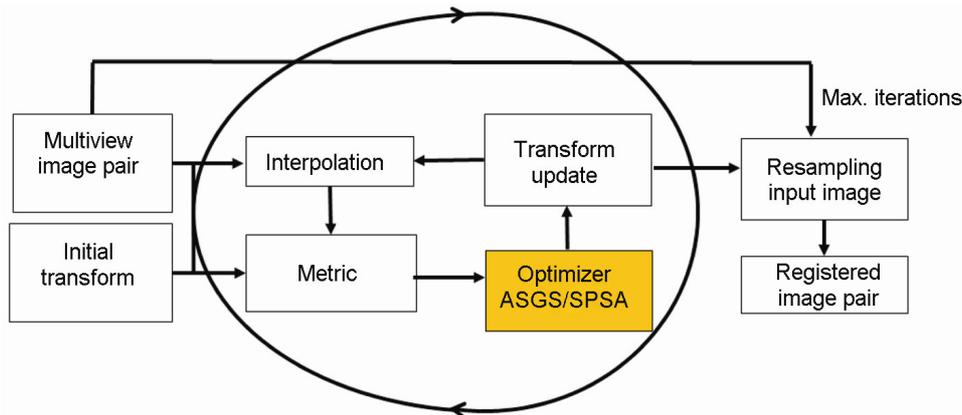
Figure 1. Optimization-based multiview satellite image registration scheme.

both techniques and check if they can handle different data conditions as well.

This article discusses technical details of the optimization mechanics employed by two different approaches, and their performance on the LISS-4 MX coregistration problem in different terrains with emphasis on the SPSA technique.

MI-based image registration

Let f_A be a reference or fixed image and f_B be a moving or floating or input image. The moving image will be registered for geometric confirmation with a transform T. Intensity-based image registration is treated as a nonlinear optimization problem.

$$\hat{\mu} = \arg\min_{\mu} C(f_A(x), f_B(T(x))), \tag{1}$$

where C is the cost function such as MI that measures similarity of the fixed image and the deformed moving image. The solution $\hat{\mu}$ is the parameter vector that minimizes the cost function. MI is an intensity-based method, i.e. grey values belonging to two different images or sub-images are directly engaged in estimating an information measure about how both images match well under the conditions set for the experiment.

The mutual information $M(A, B)$ between two random variables A and B can be defined as:

$$C(f_A(x), f_B(T(x))) = M(A, B)$$

$$= \sum_a \sum_b p_{A,B}(a,b) \log \frac{p_{A,B}(a,b)}{p_A(a) \cdot p_B(b)}. \tag{2}$$

It measures the Kullback–Leibler distance between the joint probability density function (JPDF) $p_{A,B}(a,b)$ of two random variables A and B, and the product of their

marginal PDFs, $p_A(a).p_B(b)$. To solve eq. (1), an iterative optimization procedure is employed. In every iteration k, the current transformation parameters μ_k are updated by taking a step in the search direction d_k

$$\mu_{k+1} = \mu_k - a_k d_k, \quad k = 0, 1, 2, 3, \ldots, K, \tag{3}$$

where a_k is a scalar that determines the stepsize. A wide range of optimization methods can be formulated in this way, each having different definitions of a_k and d_k. The parameter space for search is restricted only by the choice of transformation we employ as part of the image registration scheme. For an affine transform, there are six parameters defining the search space.

$$T_\mu(x) = Ax + t = \begin{bmatrix} c_0 & c_1 \\ c_2 & c_3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} t_{x1} \\ t_{x2} \end{bmatrix}, \tag{4}$$

$$\mu = [t_{x1}, t_{x2}, c_0, c_1, c_2, c_3]^T. \tag{5}$$

The optimum transformation can be obtained by an exhaustive search. This can be computationally expensive as the number of parameters increases along with sub-pixel accuracy requirement. Optimization scheme is a smart procedure to arrive at the optimum transform parameters in a guided manner. Additionally, any free-form deformation field can be modelled as a B-spline transform. It is always preferable to employ a global transform before estimating an additional free-form deformation[12]. ITK based Elastix tool has configurability to carry out image registrations with several combinations of intensity based registration components, which is used in the exercises reported here[3].

Adaptive stochastic gradient descent optimization

A common choice for the search direction in eq. (3) is the derivative of the cost function $\partial C/\partial \mu$ evaluated at the

current position defined by μ_k, commonly known as gradient descent. Under conditions, $\mu_k \to \mu^*$ in almost sure stochastic sense as $k \to \infty$. Using few samples rather than a full population, the derivative estimate may be called as stochastic estimate. If the gain sequence is made adaptive, the complete procedure may be called as ASGD optimization, which can be compared with the chosen optimizer for this study. The stochastic gradient descent method uses the following iterative scheme in place of eq. (3)

$$\mu_{k+1} = \mu_k - \gamma_k \tilde{g}_k, k = 0, 1, 2, \ldots, n, \tag{6}$$

$$\tilde{g}_k = g(\mu_k) + \varepsilon_k, \tag{7}$$

where $\tilde{g}_k$ denotes an approximation of the true derivative $g \equiv \partial C / \partial \mu$ at μ_k and ε_k is the approximation error. The estimates of g, $\tilde{g}_k$ and ε_k are normally distributed with different mean and variance measures. The scalar gain factor γ_k, the stepsize is determined by a predefined decaying function of the iteration number k. An often used choice for the gain factor is

$$\gamma_k \equiv \gamma(k) = a / (k + A)^\alpha. \tag{8}$$

The setting values for a, A and α involve the use of stochastic gradient descent (SGD) for image registration with maximum stepsize being the initial value $\gamma(0)$. With user specified setting on constants $(a > 0)$, $(A \geq 1)$ and $(0 < \alpha \leq 1)$, and the choice of $(\alpha = 1)$ gives a optimum rate of convergence when $k \to \infty$. When a is too small, the SGD method suffers from slow convergence and if it is too large, the process may become unstable[13]. The algorithm has been utilized for the adaptive stepsize in stochastic gradient descent optimization for medical image registration[7,14,15]. For the γ function, the same definition as in eq. (8) can be used. However, in ASGD, the γ function is not evaluated at the iteration number k, but at time t_k adapted depending on the inner product of the gradient $\tilde{g}_k$ and the previous gradient $\tilde{g}_{k-1}$. If the gradients in two consecutive steps point in the same direction, the inner product is positive and therefore the time is reduced, which leads to a larger stepsize $\gamma(t_{k+1})$, since γ is a monotonically decreasing function. Thus, the ASGD method implements an adaptive step-size mechanism. LISS-4 MX coregistration using ASGD and MI was exclusively dealt in the work with theoretical and experimental results[11].

Simultaneous perturbation stochastic approximation

The SPSA algorithm developed by Spall[10,16] is used here as an optimization technique in image coregistration

problem. This optimization is used where it is difficult to analytically estimate the gradient of the objective function[15]. Such situations are often encountered while computing MI between a pair of images from the probabilities measured from discrete joint histograms. SPSA is a powerful optimization technique that is based on a highly efficient gradient approximation that can be easily implemented and that relies on only two measurements of the objective function. It does not rely on explicit knowledge of the gradient of the objective function, or on measurements of this gradient as required in gradient descent optimizers such as ASGD. Thus we need to study the SPSA technique in detail. The update law mentioned in eq. (3) is more general, shared by many first-order gradient optimization methods and used by the SPSA method as well. The gradient vector for m-dimensional parameter space can be defined by

$$g_k = [g_k^1, g_k^2, \cdots g_k^m]^T. \tag{9}$$

$\{a_k\}$ in the update rule in eq. (3) is a non-negative gain sequence. Let Δ_k be a vector of p independent random variables at the kth iteration.

$$\Delta_k = [\Delta_{k1}, \Delta_{k2}, \cdots, \Delta_{kp}]^T, \tag{10}$$

where Δ_k is typically generated using the Monte Carlo method. All the components of μ_k are randomly perturbed to obtain two measurements of the objective function. Each component of the gradient vector is then formed using the ratio of the difference in these measurements and the individual components in the perturbation vector. Let $\{c_k\}$ be a sequence of positive scalars. For consecutive iterations $k \to k + 1$, measurements

$$y_{\mu_{k+c_k\Delta_k}} = C(\mu_k + c_k\Delta_k) + \varepsilon_k^{(+)}, \text{ and}$$

$$y_{\mu_{k-c_k\Delta_k}} = C(\mu_k - c_k\Delta_k) - \varepsilon_k^{(-)}$$

are done for $\mu_k \pm c_k\Delta_k$, \hfill (11)

where $\varepsilon_k^{(\pm)}$ are measurement noise terms, and implementation of SPSA often exploit a common case, when these terms are zero irrespective of the iteration number $\varepsilon_k^{(\pm)} = 0 \forall k$.

$\hat{g}_k$ is determined by a standard simultaneous perturbation (SP) form

$$\hat{g}_k(\mu_k) = \begin{bmatrix} \dfrac{C(\mu_k + c_k\Delta_k) - C(\mu_k - c_k\Delta_k)}{2c_k\Delta_{k1}} \\ \vdots \\ \dfrac{C(\mu_k + c_k\Delta_k) - C(\mu_k - c_k\Delta_k)}{2c_k\Delta_{kp}} \end{bmatrix}. \tag{12}$$

It may be noted that only two measurements of $C(.)$ are required to compute the derivative independent of transformation parameters μ_k. Simultaneous perturbation form contrasts with standard finite difference approximation taking $2p$ (or $p+1$) measurements. The following intuitive reason validates the form described above.

$$E\left[\hat{g}_k(\mu_k)|\mu_k\right] \approx g_k(\mu_k). \tag{13}$$

Essential conditions for setting up SPSA optimizer include smoothness criteria on C to be thrice continuously differentiable, Δ_k symmetric distribution around 0 and $E(\Delta_k^2) < \infty$, $E(\Delta_k^{-2}) < \infty$, gain sequences a_k and c_k as decaying functions to assume zero value neither too fast nor too slow, and noise expectation to take zero value. SPSA is a stochastic analogue to deterministic algorithms if $\hat{g}_k$ is 'on average' the same as true gradient for any transformation parameter vector μ_k. The performance of SPSA is compared to that of ASGD in the following.

Datasets and experiments

The aim of this study is to demonstrate the performance of SPSA optimizer in automatic satellite image registration tasks, especially LISS-4 MX coregistration, and also compare its performance to the standard gradient optimizer. We have chosen datasets from LISS-4 MX sensor on-board Resourcesat-2 platform[8,17]. Different spectral band data are acquired at different times within few seconds, large enough to manifest band-to-band registration differences due to various factors such as terrain relief, attitude differences, etc. Apparently, the imaging time difference may result only in fixed offsets along the track direction. However, in satellite image acquisition this is not true, owing to the fact that time difference of a few seconds has to account for satellite position, velocity, the attitude profiles being continuously updated or changing and terrain relief. However, in an image-to-image registration model philosophy, these issues can be tackled by using rather different techniques than modelling as satellite sensor navigation issues[18]. Feature and navigation model-based satellite image registration exercise encounters choice of navigation model, correction levels and object condition and earth model. Earth is modelled as an ellipsoid and the apparent topography is dealt using elevation data above ellipsoid[11]. This approach makes use of neither the navigation model nor DEM for registration, but is able to model the deformation fields represented by affine and B-spline transformations together. Typical terrain conditions such as urban, plain, undulating, cloudy scenes are also important issues while addressing registration performance limits. To study few such aspects, representative datasets were chosen (Table 1) belonging to various dates of imaging with coverage in urban, ocean, snow and cloudy situations (Figure 2). The sub-image regions marked in the figure are identified with urban, land–ocean, terrain and cloud patches. Three spectral bands are identified as Band 2 (B2), Band 3 (B3), Band 4 (B4) corresponding to wavelengths 420–470 nm, 520–580 nm and 620–680 nm respectively.

It is important to understand how the chosen optimization schemes view the objective function terrain. Figure 3 shows the similarity metric surface encountered by the optimization procedures against simple translations for the urban features-dominated images shown above. The contours of the metric surface projected sideways indicate how sharp both the optimizers are while finding the minima, and bottom plane contours show the smoothness, continuity and regularity of the search space.

Figure 4 shows MI, its derivative and the displacement to register the images and it indicates that ASGD reaches optimum in fewer iterations than SPSA. Nevertheless, they match at the final iterations. It is important to note that, without any initial transform ASGD has performed in lesser number of iterations than SPSA, but with an initial transform (18.0 pixel shift in sample direction) closer to the actual solution, ASGD as well as SPSA exhibit similar performance. These exercises confirm that ASGD and SPSA are similar techniques; the latter may require more iterations than the former when the initial solution is far away from the final solution.

Results and analysis

Optimization mechanics is efficient when the initial condition is closer to the final or optimum solution at the final scale of data which the user wants. When the optimization for registration is performed directly at the final resolution/scale of the data, the number of iterations has to be sufficiently large, which is difficult to estimate. This problem is overcome by setting up multiresolution optimization procedure; i.e. image pair is represented in a multiresolution pyramid and registration starts with the lowest pyramid bringing the advantage of capturing the displacement at a scaled-down level, which is progressively refined at finer resolution. This has an additional advantage of avoiding local minima. The full-size image (12,000 × 17,000 pixels) registration was done using affine transformation for global deformation and B-spline field for local deformation together. Four levels of image pyramids from one eighth to original pixel resolution were set-up with the above-mentioned choices of transformation along with ASGD and SPSA optimizers. While we perform image registration with global and local deformation models represented by affine and B-spline transformations, evaluation of the registration performance is reported as mean residuals along two image axes, namely line and pixel directions (translations) by independent exercises. The same image registration model can be used to measure the residuals by repeating the model execution

Table 1. Coregistration performance measured by the registration algorithm

Dataset	Terrain	Unregistered		ASGD		SPSA	
		Scan	Pixel	Scan	Pixel	Scan	Pixel
30 May 2012 C, 107/58	Urban-2	1.869	21.101	0.057	−0.028	0.044	−0.004
	Urban-4	−1.134	−13.98	0.108	−0.03	0.106	−0.014
8 November 2011 B, 95/62	Ocean-2	−0.110	22.152	0.225	0.0781	0.011	0.012
	Ocean-4	0.862	−14.477	−1.064	−0.835	−0.876	−0.900
3 November 2011 D, 94/46	Terrain-2	7.391	20.096	−0.056	−0.04	−0.033	0.014
	Terrain-4	−6.216	−13.36	0.0153	0.019	0.011	−0.030
	Cloud-2	11.312	20.272	0.206	−0.132	0.228	0.288
	Cloud-4	−8.787	−13.499	−0.11	0.072	−0.207	0.152

ASGD, Adaptive stochastic gradient descent; SPSA, Simultaneous perturbation stochastic approximation.
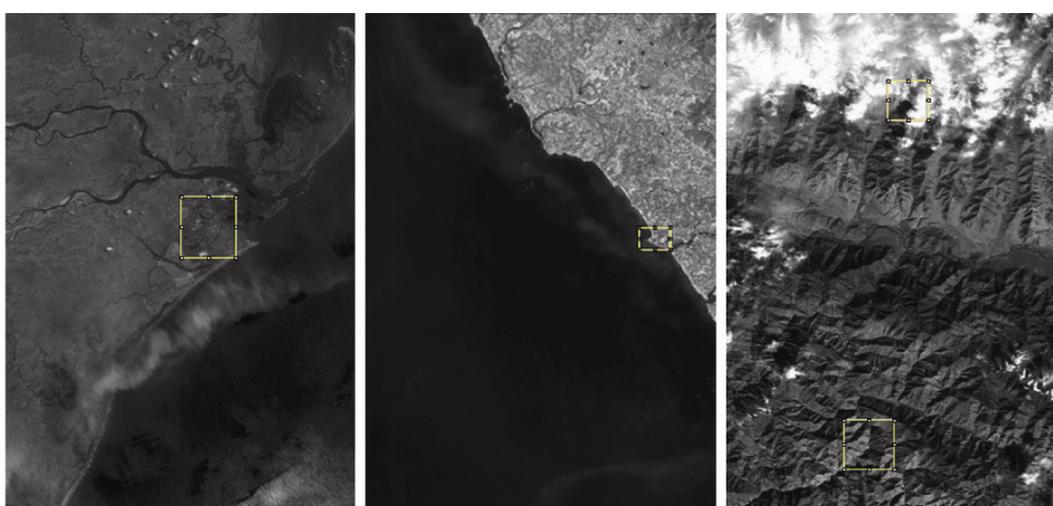
Figure 2. LISS-4 B3 images acquired on 30 May 2012, 8 November 2011 and 3 November 2011 corresponding to urban, land–ocean and snow–cloud coverages marked by boxes.
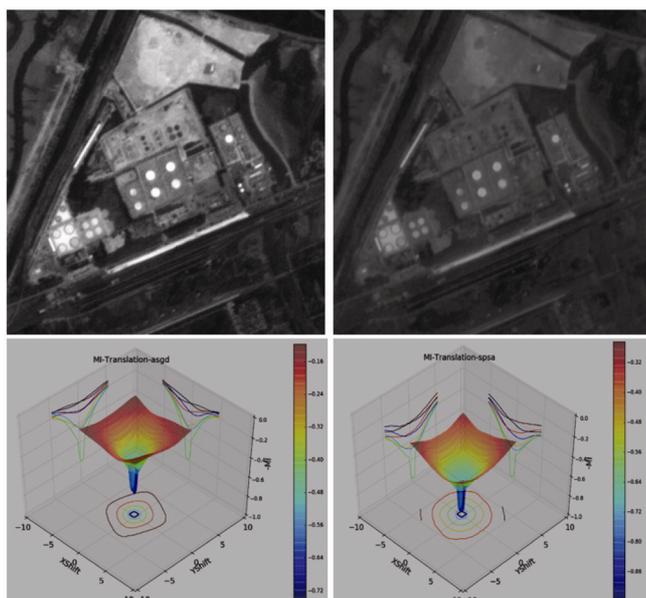
Figure 3. Urban subimages of B3 and B2 (first row left and right) data with mutual information search space.

once more, this time with registered set only for taking a performance measure. This can be ascertained for sub-images marked over full images as shown in Figure 2 and in Table 1. Urban2/4 indicates that image region belongs to urban feature and the registration task attempted was Band 2/Band 4 data to Band 3 coregistration.

Table 1 shows translations estimated by the image registration model. Large translations are seen in the unregistered input images for different datasets, and translations alone may not be sufficient to overlay images precisely, as stated earlier. However, the estimated translations for the registered cases are of the same order using both the optimizers, indicating that their performances are comparable. An independent assessment of registration performance is required which is not biased by the same approach used to achieve the results. Fourier domain-based phase correlation is a robust and global shift estimator for overlapping images[19]. Table 2 also shows residual translations but after employing affine and B-spline transform models together and estimated by phase correlation technique, another independent way of estimating
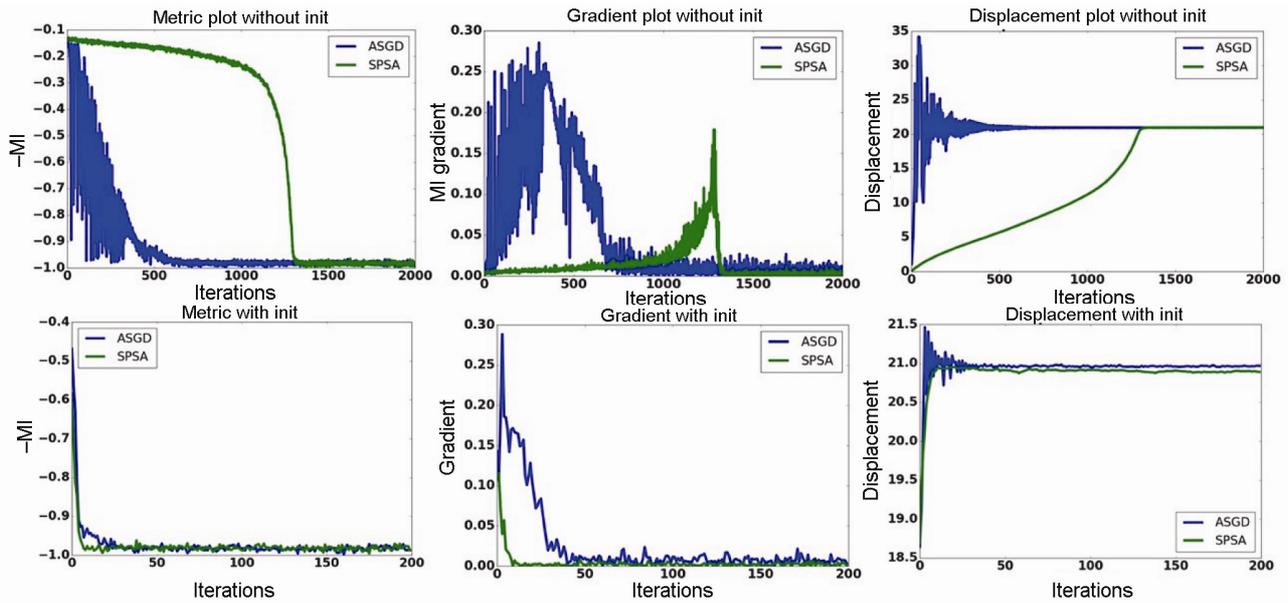
Figure 4. ASGD and SPSA optimization relative performance with and without initial transform set-up for urban region.
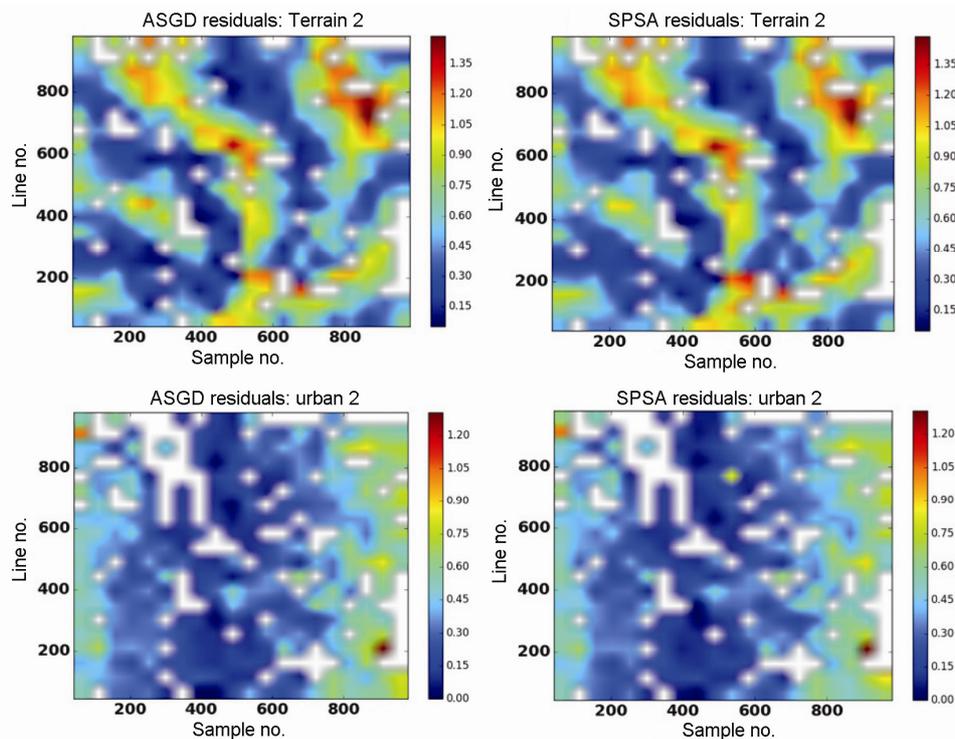
Figure 5. Residuals plot map using ASGD (left column) and SPSA (right column) registered cases with white-space representation for no value spots.

the residual translations, and exhibits similar order of accuracies.

Figure 5 shows registration performance using another independent exercise of automatic landmark feature matching by spatial domain image correlation sensitive up to 0.1 pixel and calculating residuals thereafter in scan and pixel directions individually. The residuals (bounded within ±0.5 pixels) plot map indicates that the performance scale is similar for both the gradient-based optimizer techniques and either of them can be used in automatic satellite image registration exercises. The continuous residual map is constructed from irregular grid data by 2D binning and interpolation. The white legend indicates there are no matching data available for interpolation. This could

Table 2. Coregistration performance global score by phase correlation

	Unregistered		ASGD		SPSA	
Dataset	Scan	Pixel	Scan	Pixel	Scan	Pixel
Urban-2	1.682	21.077	0.003	−0.065	−0.010	−0.032
Urban-4	−1.245	−13.825	0.014	−0.002	0.021	0.020
Ocean-2	−0.129	22.057	0.024	−0.181	0.032	−0.082
Ocean-4	−0.781	−14.983	−0.034	−0.149	−0.482	−0.421
Terrain-2	7.477	20.073	−0.007	−0.023	−0.031	−0.069
Terrain-4	−6.226	−13.279	0.024	0.054	0.027	0.093
Cloud-2	10.631	20.716	−0.120	0.108	−0.194	0.152
Cloud-4	−8.312	−13.743	0.122	−0.084	0.208	−0.139
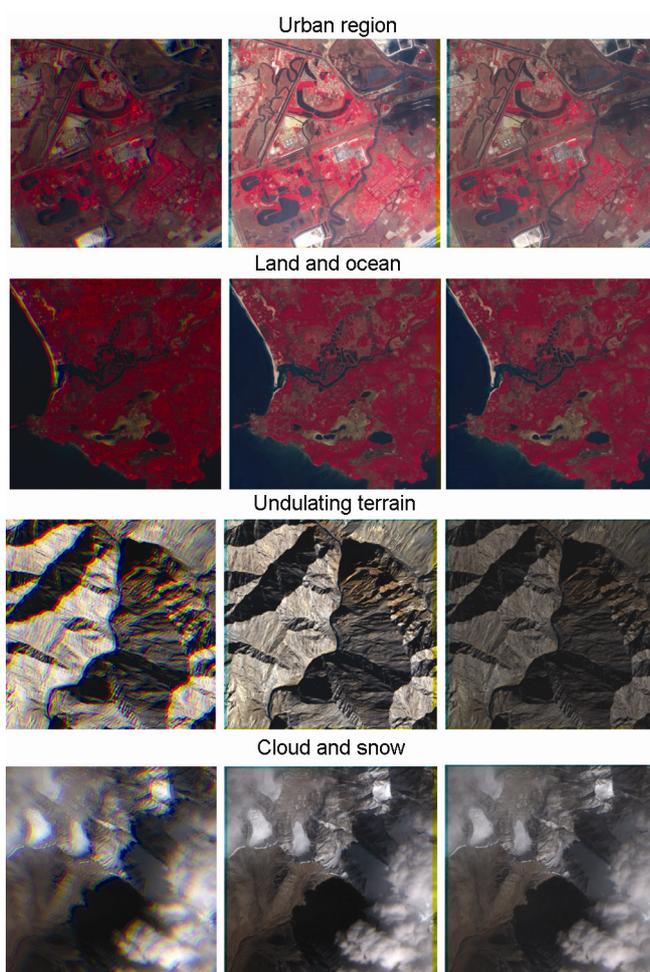
Figure 6. Colour composites of different terrain data for unregistered, ASGD-registered and SPSA-registered sets.

be an efficient way of assessing the performance of the registration across image regions instead of calculating RMSE figures as generally followed.

Colour composite displays are important for visual scrutiny of coregistered multispectral datasets. The unregistered multispectral images show colour artifacts in the overlapping features, more pronounced in the con-trasting image regions (Figure 6). The registered colour composites do not show any such artifacts in either of the registration approaches. The difference in imaging time in LISS-4 multispectral data is of the order of only 1–1.4 sec relative to the central band 3. Thus there is very little cloud movement during that time duration and snow is a static feature in that time difference, and hence they are registered. If there are moving clouds, they exhibit colour fringes indicating cloud motion. The local distortions are represented by a B-spline field to take care free-form deformations. Registration of static features is assured by doing repeated stochastic sampling and the rest of procedures in a number of iterations. It is to be understood that moving pixel population is less compared to unchanging background features. Therefore the moving samples do not influence the cost function derivative so much compared to the static or permanent features. If data occlusion is more (say >50%), then the registration is unstable as found in our experiments.

The advantage in SPSA in comparison to ASGD is that no exclusive gradient needs to be estimated; only cost functions are evaluated and perturbed to estimate the same. Otherwise, they belong to similar stochastic optimization regimes, however, SPSA may require additional iterations as mentioned earlier.

Conclusion

Gradient-based optimizers have been studied in intensity-based satellite image registration using ASGD and SPSA. These exercises confirm that the performance scales are similar to both the optimizers; however, SPSA has the advantage of not estimating the gradient analytically as in the case of ASGD. SPSA can also be used when direct estimation of derivatives is not possible or noisy. SPSA is a good optimizer choice for any image registration task aiming for subpixel accuracy performance. This can be validated for different terrains data sets. Muti-resolution pyramid, random sampling for cost function estimation to overcome noisy input data should be in place for good performance. For highly undulating terrains, in addition to global transformations such as affine, B-spline

represented deformation models are necessary to ensure acceptable satellite image registration performances. The performance of registration is better assessed by a continuous map rather than projecting a single RMSE value.

1. Maes, F., Collignon, A., Vandermeulen, D., Marchal, G. and Suetens, P., Multimodality image registration by maximization of mutual information. *IEEE Trans. Med. Imaging*, 1997, **16**, 187–198.
2. Thévenaz, P. and Unser, M., A pyramid approach to sub-pixel image fusion based on mutual information. In Proc. IEEE Int. Conf. Image Processing, Lausanne, Switzerland, 16–19 September 1996, pp. 265–268.
3. Klein, S., Staring, M., Murphy, K., Viergever, M. A. and Pluim, J. P. W., Elastix: a toolbox for intensity-based medical image registration. *IEEE Trans. Med. Imaging*, 2010, **29**, 196–205.
4. Mattes, D., Haynor, D. R., Vesselle, H., Lewellen, T. K. and Eubank, W., PET-CT image registration in the chest using free-form deformations. *IEEE Trans. Med. Imaging*, 2003, **22**, 120–128.
5. Unser, M., Splines: a perfect fit for signal and image processing. *IEEE Signal Process. Mag.*, 1999, **16**, 22–38.
6. Goshtasby, A. A., Registration of image with geometric distortion. *IEEE Trans. Geosci. Remote Sensing*, 1988, **26**, 60–64.
7. Klein, S., Staring, M. and Pluim, J. P. W., Evaluation of optimization methods for nonrigid medical image registration using mutual information and B-splines. *IEEE Trans. Image Process.*, 2007, **16**, 2879–2890.
8. Manthira Moorthi, S., Gambhir, R. K., Ramakrishnan, R. and Sivakumar, R., Performance study of optimization methods for intensity-based automatic satellite image registration. *Int. J. Imaging Robotics*, 2012, **8**, 101–110.
9. Cole-Rhodes, A. A., Johnson, K. L., LeMoigne, J. and Zavorin, I., Multiresolution registration of remote sensing imagery by optimization of mutual information using a stochastic gradient. *IEEE Trans. Image Process.*, 2003, **12**, 1495–1511.
10. Li, Qi, Sato, I. and Murakami, Y., Simultaneous perturbation stochastic approximation algorithm for automated image registration optimization. In Geoscience Remote Sensing Symposium, IEEE International Symposium on Geoscience and Remote Sensing, Denver, CO, USA, 2006, pp. 184–187.
11. Manthira Moorthi, S., Dhar, D. and Sivakumar, R., Coregistration of LISS-4 multispectral band data using mutual information based stochastic gradient descent optimization, *Curr. Sci.*, 2017, **113**, 877–888.
12. Rueckert, D., Sonoda, L. I., Hayes, C., Hill, D. L. G., Leach, M. O. and Hawkes, D. J., Nonrigid registration using free-form deformations: application to breast MR images. *IEEE Trans. Med. Imaging*, 1999, **18**, 712–721.
13. Robbins, H. and Monro, S., A stochastic approximation method. *Ann. Math. Stat.*, 1951, **22**, 400–407.
14. Plakhov, A. and Cruz, P., A stochastic approximation algorithm with step size adaptation. *J. Math. Sci.*, 2004, **120**, 964–973.
15. Klein, S., Pluim, J. P. W., Staring, M. and Viergever, M. A., Adaptive stochastic gradient descent optimization for image registration. *Int. J. Comput. Vis.*, 2009, **81**, 227–239.
16. Spall, J. C., Multivariate stochastic approximation using a simultaneous perturbation gradient approximation. *IEEE Trans. Autom. Control*, 1992, **37**, 332–341.
17. Manthira Moorthi, S., Kayal, R., Ramakrishnan, R. and Srivastava, P. K., Resourcesat-1 LISS-4 MX bands on ground co-registration by in-flight calibration and attitude refinement. *Int. J. Appl. Earth Obs. Geoinf.*, 2008, **10**, 140–146.
18. Theiler, J. P., Galbraith, A. E., Pope, P. A., Ramsey, K. A. and Szymanski, J. J., Automated coregistration of MTI spectral bands. *Proc. SPIE*, 2002, **4725**, 314–327.
19. Mark, A. G., Sub-pixel registration assessment of multispectral imagery. *Proc. SPIE, Imaging Spectrom. XI*, 2006, **6302**, 1–12.