

RETRIEVAL OF DEFORMATION FIELDS BY USING STOCHASTIC MUTUAL INFORMATION BASED OPTIMIZATION IN AUTOMATIC REGISTRATION OF SATELLITE IMAGES

Subbiah Manthira Moorthi¹ and Ramamoorthy Sivakumar²

¹ODPD/SIPG, Space Applications Centre, Indian Space Research Organisation, Gujarat, India

²Department of Civil Engineering, SRM Institute of Science and Technology, India

Abstract

Modeling and retrieving the transform parameters that characterize the underlying deformation field is the main crux of the problem in automatic image registration domain which involves employing a similarity measure in an image pair and a robust model estimator. Model estimators can be either a least square fit or an optimization method which finds minimum of a cost function. In this work, a stochastic mutual information based adaptive gradient descent optimizer is proposed in which transforms such as translation, affine and free form deformations are accurately retrieved in the process of image registration and only a percentage of population of intensities is used to estimate mutual information without losing accuracy in a stochastic way. Better than one tenth of a pixel accuracy is achieved in image registration by retrieving different geometric transformations accurately.

Keywords:

Mutual Information, Image Registration, Optimization, Deformation, Transforms

1. INTRODUCTION

Satellite images of same geographic extent can be obtained in different view angles or any other conditions related to sensor mechanism, spectral channels and modalities. Remote sensing data users invariably prefer registered multi date data sets for analyzing further for typical applications such as change detection studies. In image registration the feature matching criteria or similarity measure decides the final geometric fidelity of the input image to the reference image. Correlation measure assumes a linear photometric model of intensities between the input and the reference images, which becomes a limiting factor when it comes to handling multi modal images, where the relationship need not fall into a linear model [1]. Correlation measure does not even perform matching efficiently between visible and infra-red channels of satellite remote sensing data sets.

A classic survey article by Brown [2] classifies image registration methods based on the different combinations of the components used for image registration namely, a feature space, a search space, a search strategy, and a similarity metric. However, the emphasis was more on the feature based and correlation technique based methods as intensity and Mutual Information (MI) based methods were yet to become prominent in this field. Feature based image registration methods are elaborately discussed also by Zitova et al. [3] and Zhen and Yun [4]. We bring out the details of employing mutual information based metric and cost function in an optimization framework to achieve automatic satellite image registration, known as an intensity based approach.

MI is related to intensity based method, as the gray 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 registration [5]. Estimation of MI does not require any preprocessing step of extracting feature sets which enable automatic image registration fully. It is known that MI base image registration was introduced independently by two different research groups lead by Viola and Wells [6] and Collignon [1] in the area of medical image registration. Many researchers extended MI based techniques to solve medical image registrations that used multimodal medical images [6]-[8]. It was only after year 2000, MI based techniques made entry into image registration scenario of remote sensing domain. Since then there has been a steady progress made by several researchers Lemoigne [9], Cole-Rhodes [10], and Chen [11]. However, characteristics of remote sensing images are so varied, and additional issues like cloud presence, presence of large homogenous features such as ocean, snow cover etc. pose further problems to look for different techniques to automate satellite based image registration tasks. It is worth investigating image registration problems drawn from different satellite borne sensor images in a completely automatic manner by retrieving the deformation fields modeled by the transform choice.

MI measure produces sharper peaks under rotation or translation, than the cross correlation. Moreover, sharper peaks are also produced at the lowest resolution of the images in the resolution pyramids. This indicates that MI can produce more accurate results than correlation in a multi resolution registration scheme as shown by Thevenaz and Unser [12]. Implementation issues which influence the estimation of MI measure and the accuracy of the image registration results include the estimation of the probability density functions (PDFs), interpolation methods for the PDFs and/or the images, and optimization methods. Optimization methods demand estimation of gradient of the cost functions employed such as MI. Core theoretical concepts used in the paper are covered in the next section.

2. THEORETICAL BASIS FOR MUTUAL INFORMATION

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

$$M(A,B) = \sum_a \sum_b p_{A,B}(a,b) \log \frac{p_{A,B}(a,b)}{P_A(a)P_B(b)} \quad (1)$$

It measures the Kullback-Leibler distance [5] between the joint PDF $P_{A,B}(a,b)$ of two random variables A and B and the product of their marginal PDFs, $P_A(a)$, $P_B(b)$. Thus, it measures

the distance of the joint PDF of the random variables A and B from that associated with complete independence of A and B , implying that MI is zero if and only if the two random variables are independent. Thus MI is a measure of the statistical dependence between two datasets indicating the amount of information that random variable A contains about random variable B , and vice versa. You can also express MI in terms of entropies as follows,

$$M(A, B) = H(a) + H(b) - H(a, b) \quad (2)$$

where

$$H(a) = -\sum_a p(a) \log p(a)$$

$$H(b) = -\sum_b p(b) \log p(b)$$

$$H(a, b) = -\sum_a \sum_b p(a, b) \log p(a, b)$$

Let f_A and f_B be the reference and floating images engaged in registration. The joint probability distribution, given a geometric transformation T , is estimated by first computing a two-dimensional frequency histogram $H(A, B, T)$ of the gray value combinations $(f_A(x), f_B(T(x)))$ for every grid point x of the floating image f_B that falls inside the region of overlap of the two images. The entries in the histogram are then divided by the total number of entries to create a probability distribution. The grey value $f_B(T(x))$ is estimated, by interpolation using neighboring pixels in the mapped location. The histogram may be normalized to give an approximation of the probability distribution function (PDF) of intensities, i.e.

$$p_{A,B}(a, b) = \frac{H(a, b)}{N_x} \quad (3)$$

where, N_x is the number of samples in the histogram. The marginal probability distributions for f_A and f_B are calculated by summation of either the rows or the columns of the joint histogram. The PDF's P_A and P_B are easily obtained from the joint PDF, since $P_A = \sum_a p_{A,B}$ and $P_B = \sum_b p_{A,B}$. Treating a and b as discrete variables (or indices) indicating the finite bin size of the histogram H from which PDF is derived. MI is not invariant to the bin size Δi , which limits its bounds, as does the number of sample points:

$$I \leq \log \left(\min \left(\frac{a_{mx}}{\Delta_i}, \frac{a_{mx}}{\Delta_i}, k_{mx} N_x \right) \right)$$

where, k_{mx} indicates the number of histogram bins populated per sample.

2.1 MI USING PARZEN WINDOWING

The approach presented here uses MI as the cost function in an optimization framework for registering satellite images and it will become further clear when we discuss about the optimizer, that the estimation of derivative of MI is essential to setup the optimizer mechanics. The discrete estimation of MI does not ensure derivatives, and hence density estimates for probability distributions based on so called Parzen windowing is employed [12]. This will support to represent MI as an analytical function to estimate derivative of the same. In this connection, let us define

the joint histogram in terms of two window functions $\omega()$, which act as membership functions:

$$\#(A, B) = \sum_x \omega \left(a - \frac{f_A}{\Delta_i} \right) \omega \left(b - \frac{f_B(T(x))}{\Delta_i} \right) \quad (4)$$

where, each sample taken from f_B and f_A is added to one histogram bin:

$$\omega(\varepsilon) = \beta_0^-(\varepsilon) \equiv \begin{cases} 1 & 0 \leq \varepsilon < 1 \\ \text{otherwise} & \end{cases} \quad (5)$$

This kind of sampling is referred to as standard sampling. The $\beta()$ function in the above equation comes from the B-spline family of functions with several useful properties, a brief description of which is given here. A more detailed description of B-spline functions and their numerical computation is given by Unser [13], but we will cover few critical properties. Firstly, the sum of a B-spline function for all integral distances from a real value is one, i.e. it has a portion of unity. This means that no renormalization is required when histogramming. Secondly, the integral of a B-spline is one. Thirdly, order n B-splines are the convolution of any set of B-splines whose order sums to n . Finally, the derivative of an order n B-spline is a function of two order $n-1$ B-splines. These properties are summarized below.

$$\sum_{a \in \mathbb{Z}} \beta(x+a) = 1, x \in \mathbb{R} \quad (6)$$

$$\int_{x \in \mathbb{R}} \beta(x) \equiv 1 \quad (7)$$

$$\beta^n \equiv (\beta^{n-1} \times \beta^0)(x) \equiv \int_{-\infty}^{\infty} \beta^{n-1}(x) \beta^0(x-t) dt, n > 0 \quad (8)$$

$$\frac{\partial \beta^n}{\partial x} \equiv \beta^{n-1} \left(x + \frac{1}{2} \right) - \beta^{n-1} \left(x - \frac{1}{2} \right) \quad (9)$$

where β^0 is a unit square pulse, as given below

$$\beta^0(x) = \frac{1}{2} \left(\text{sign} \left(x + \frac{1}{2} \right) \text{sign} \left(x - \frac{1}{2} \right) \right) \quad (10)$$

where,

$$\text{sign}(x) = \begin{cases} -1 & x < 0 \\ 0 & x = 0 \\ +1 & x > 0 \end{cases} \quad (11)$$

B-splines are used for Parzen windowing for density estimates as it satisfies the required conditions such as function with unit integral and partition of unity property etc. The joint histogram is redefined and hence a notation change is introduced. The joint histogram is given by the following expression in terms of two B-spline window functions $\omega()$, which act as membership functions

$$h(f, m, v) = \frac{1}{\varepsilon_B \varepsilon_A} \sum_{x_i \in V} \omega \left(\frac{f - f_A(x_i)}{\varepsilon_A} \right) \omega \left(\frac{m - f_B(T(x_i; v))}{\varepsilon_B} \right) \quad (12)$$

where, f, m are grey scale values and $\varepsilon_A, \varepsilon_B$ are the window scaling factors, V the overlap region and v the transform parameters. The joint probability is now expressed as

$$\tilde{P}(f, m, v) = \alpha(v) h(f, m, v) \quad (13)$$

where, $\alpha(v) = \frac{1}{\sum_{f,m} h(f,m,v)}$. Recall MI expression as

$$M = \sum_{f=1}^{L_f} \sum_{m=1}^{L_m} \hat{P}(f,m,v) \log \left(\frac{\hat{P}(f,m,v)}{\hat{P}_A(f) \hat{P}_B(m)} \right) \quad (14)$$

where, L_f and L_m are the sizes of the samples drawn from reference and floating images.

2.2 STOCHASTIC ESTIMATES OF HISTOGRAM, PDF AND MI

PDF and MI estimates can be done either with full population of intensities or selected subset of samples, the later one is called as stochastic estimate. Stochastic estimates of joint histogram, joint probability and MI are compared in Fig.2 and Fig.3 for the image shown in Fig.1. The stochastic estimate is brought by only a sub set of entire intensity samples which are randomly chosen from the image regions set within the image to be registered. In Fig.2 and Fig.3 joint histogram, and joint PDF estimated through various sample sizes are depicted. It shows that, the entire population of intensities is not required, only a sub set of about 10% of samples (learned from our experiments) would be enough to achieve close to the deterministic estimates and it would be advantageous to provide a robust estimate even in the case of data occlusions such as cloud presence in satellite images. Joint histogram data are used for estimating MI values which are tabulated in Table.1.

It can be verified that 10% (~10000 samples) and above samples are close enough to final deterministic MI value of 9.9237 using the entire pixel population.



Fig.1. Satellite image used (LISS-3, 2nd Jan, 2011, path or row 093/56), sub image and shifted version for the inscribed box

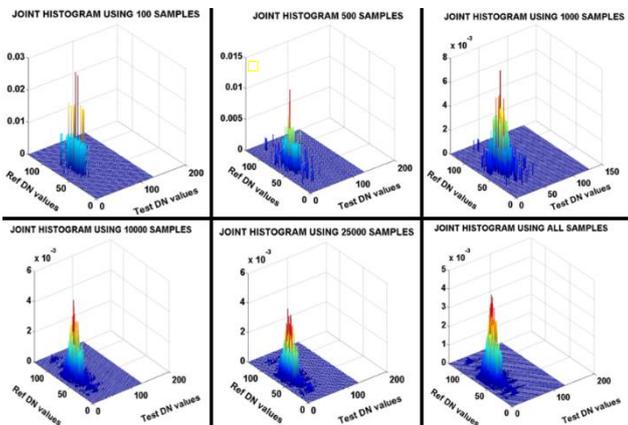


Fig.2. Joint histogram for images in Fig.1 with different sample sizes

The similarity measure surface profile with respect to the transform parameters should be smooth enough for search and optimization to converge smoothly otherwise optimization procedures may face problems due to local minima. But smoothness is not the only criteria but it should also be sharply falling or rising at the extremum point. The Fig.4 portrays MI surface between an image and its transformed version with translations steps in both the image directions. For comparison purpose the Normalized Cross Correlation (NCC) surface is also shown. NCC surface may be smoother, but negative value of MI displays a distinct and a sharp valley which would be useful in arriving at the final transform in a robust manner.

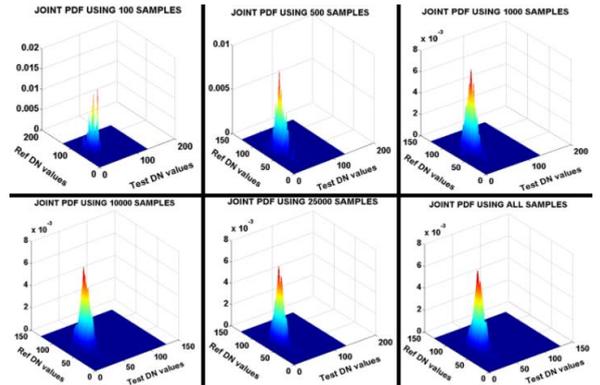


Fig.3. PDF for images in Fig.1(b) with different sample sizes

Table.1. Stochastic estimate of MI from varying sample size

No. of samples	Mutual Information Value	Time in seconds
100	8.9099	11.12
500	9.5279	11.32
1000	9.6603	11.79
2000	9.7472	12.43
5000	9.8500	13.52
10000	9.8891	15.94
25000	9.9044	23.44
All samples	9.9237	59.82

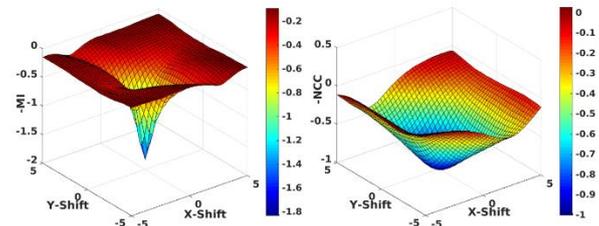


Fig.4. MI and NCC metric surface w.r.t 2D translations

3. IMAGE REPRESENTATION, DERIVATIVE AND FREE FORM DISTORTIONS

Let us recollect that, MI was estimated using B-spline Parzen windowing technique. We also remind the readers that any image

can be modeled with separable B-spline basis functions discussed earlier as

$$f(x) = \sum_{x_i \in V} c(x_i) \beta^n(x - x_i) \tag{15}$$

where, $c(x_i)$ are the constants determined from the image data itself, with x_i defined over $v \subset \mathbb{R}^2$. Image gradients are calculated as

$$\frac{\partial f(x)}{\partial x} = \sum_i c_i \left(\frac{\partial \beta^3}{\partial u} \Big|_{u=x-x_i} \beta^3(y - y_i) \right) \tag{16}$$

with $\frac{\partial f(x)}{\partial y}$ analogous.

The derivative of the similarity measure usually involves computation of the spatial derivative of the moving or test image: $\frac{\partial I_{f_B}}{\partial x}$, and the derivative of the transformation to its parameters: $\frac{\partial I_{f_B}}{\partial x}$. In the derivative to the transformation parameters v is referred to as Jacobian in technical parlance. The derivative of MI may now be found by applying the product and chain rules and collecting the terms:

$$\frac{\partial M}{\partial v_i} = \sum_{j=1}^{L_f} \sum_{m=1}^{L_m} \frac{\partial \hat{P}(f, m, v)}{\partial v_i} \log \left(\frac{\hat{P}(f, m, v)}{\hat{P}_B(m)} \right) \tag{17}$$

Because the marginal PDF of the fixed image A has no influence on the derivative a simple zero order B-spline (NN interpolation) is used to estimate this PDF. Using a cubic B-spline Parzen window for the source or reference image B and the transformation independence of $\hat{P}_A(f)$ and α in Eq.(13), the joint PDF can be expressed as:

$$p(f, m; v) = \frac{\alpha}{\epsilon_B \epsilon_A} \sum_{x_i \in V} \beta^3 \left(\frac{m - f_B(T(x_i; v))}{\epsilon_B} \right) \beta^0 \left(\frac{f - f_A(x_i)}{\epsilon_A} \right) \tag{18}$$

$$\frac{\partial p(f, m; v)}{\partial v} = \frac{1}{\epsilon_A N} \sum_{i=1}^N \beta^0 \left(\frac{f - f_A(x_i)}{\epsilon_A} \right) * \left(\frac{\partial \beta^3(u)}{\partial u} \Big|_{u=\frac{m - f_B(T(x_i; v))}{\epsilon_B}} \right) \left(\frac{-\partial f_B(t)}{\partial t} \Big|_{t=T(x; v)} \right)^T \frac{\partial T(x; v)}{\partial v_i} \tag{19}$$

3.1 TRANSFORMATIONS

The image to image geometric relationship is decided by the transformation model employed such as a single global affine transform for the entire image which incorporates a translation by default apart from scaling, rotation and shear components described by a matrix A in the equation below.

$$T_v(x) = Ax + t = \begin{bmatrix} c_0 & c_1 \\ c_2 & c_3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} t_{x1} \\ t_{x2} \end{bmatrix} \tag{20}$$

where $v = [c_0, c_1, c_2, c_3, t_{x1}, t_{x2}]^T$ is the coefficient vector engaged in the above equation.

B-spline based free form deformation can be modeled with the domain of image as,

$$\Omega = \{(x, y) \mid 0 \leq x < X, 0 \leq y < Y\}.$$

Let ϕ denote a $n_x \times n_y$ mesh of control points $\phi(i, j)$ with uniform spacing δ . Then, the FFD can be written as product of the familiar 1-D Cubic B-splines

$$T_{local}(x, y) = \sum_{\ell=0}^3 \sum_{m=0}^3 \mathcal{B}^\ell(u) \mathcal{B}^m(v) \phi_{i+\ell, j+m} \tag{21}$$

where

$$i = \lfloor x/n_x \rfloor - 1,$$

$$j = \lfloor y/n_{xy} \rfloor - 1,$$

$$u = x/n_x - \lfloor x/n_x \rfloor,$$

$v = y/n_y - \lfloor y/n_y \rfloor$ and \mathcal{B}^ℓ represents the ℓ th basis function of the B-spline.

$$\mathcal{B}^0(u) = (1-u)^3 / 6$$

$$\mathcal{B}^1(u) = (3u^3 - 6u^2 + 4) / 6 \tag{22}$$

$$\mathcal{B}^2(u) = (-3u^3 + 3u^2 + 3u + 1) / 6$$

$$\mathcal{B}^3(u) = u^3 / 6$$

The basis functions of cubic B-splines have a limited support, i.e. changing control point affects the transformation only in the local.

4. AN OPTIMIZER FRAMEWORK

Automatic estimation of the transformation model that warps the input or moving image to reference image geometric reference is the sole aim of image registration task. This estimation cannot be arrived in a single shot, usually achieved in many iterations whether the approach is manual or automatic. The iterative framework chosen should be intelligent enough to guide the estimation of transformation model based on the similarity metric employed. The iterative computational setup is called as an optimizer which plays an important role for image registration tasks as it uses the metric criterion as the objective function along with the transform parameters. Transform model parameters are updated in every optimization loop driven by two parameters namely the step length and direction.

The final aim of image registration problem is to arrive at an appropriate transform model that relates the images in supplied resolution rather than at any other spatial resolution. However, there cannot be any restrictions on how it is arrived at a particular resolution. This can be achieved by employing multi resolution strategy using image pyramids constitute a sequence of reduced resolution images created from the input images. In general, this coarse-to-fine hierarchical strategy applies to the usual registration methods, and it starts with the reference and input images on a coarse resolution [12].

Master or fixed image denoted by $f_A(x)$ serve as the geometric reference for the moving or slave image $f_B(x)$ to be registered with employing a transformation $T(x,v)$. A satellite image refers to a 2-dimensional light intensity function, gives a value at position $x \in \mathbb{R}^2$.

$$\hat{v} = \arg \min_v O(f_A, f_B \circ T) \quad (23)$$

where, O is the cost function that measures the similarity of the master image and the deformed slave image [12]. MI always yields positive values, and so for our minimization problem, negative value of the MI is used. The solution $\hat{v} \in R^P$ is the transform parameter vector that minimizes that cost function. To solve Eq.(2), an iterative optimization procedure is employed with gradient descent mechanics. In every iteration k , the current transformation parameters are updated by taking a step in some search direction d_k with a step size a_k .

$$v_{k+1} = v_k - a_k d_k, k = 0, 1, 2, \dots, K \quad (24)$$

The first derivative of MI, d_k is used to setup the optimizer mechanics. A stochastic gradient descent (SGD) optimizer approximates the gradient of the cost function and gives economy in computations explained with details in Klein works [15]-[17] and we have elaborated the same in earlier section also.

Further, an adaptive SGD method uses the following iterative scheme:

$$v_{k+1} = v_k - \beta_k \tilde{g}_k, k = 0, 1, 2, \dots, K \quad (25)$$

$$\tilde{g}_k = g(v_k) + \varepsilon_k \quad (26)$$

where \tilde{g}_k denotes an approximation of the true derivative $g \equiv \partial C / \partial v$ at v_k , and ε_k is the approximation error. The scalar gain factor β_k , the step size, is determined by a predefined decaying function of the iteration number k . An often used choice is:

$$\beta_k \equiv \beta(k) = a / (k + A)^\alpha \quad (27)$$

Setting values for parameters a , A , and α in Eq.(17) complicates the usage of SGD for image registration. Specification of constants ($a > 0$), ($A \geq 1$) and ($0 < \alpha \leq 1$), choice of ($\alpha = 1$) gives an optimum rate of convergence when $k \rightarrow \infty$. If the cost function gradients in two consecutive steps point in the same direction, the inner product will be made positive, and therefore the time is reduced, which leads to a larger step size. A clever choice of sampling intensities is imperative in this framework to estimate MI and its derivative. Instead of engaging the whole image intensities, a sub set can be employed choosing fresh sample set in every iteration to reduce computations yet achieve convergence in this minimization problem which has been resolved in many research findings [18]. The Fig.5 illustrates the gradient descent optimization process in registering with a designed translation transform of (5.675, 3.656) pixels and retrieved the translation by the registration process (5.732, 3.587) in 500 iterations for the image shown in Fig.1. In every iteration, MI, its gradient, the step size and the translation estimated are plotted in Fig.5 shows how the method retrieves the translation parameters with no initial transform parameters.

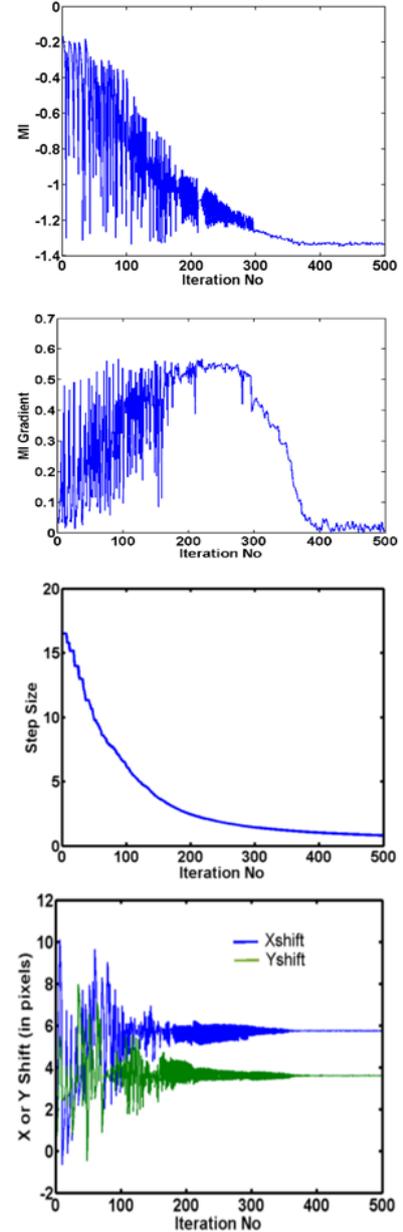


Fig.5. Gradient Descent optimization MI, MI Gradient, Step size and estimated Translations vs Iteration

5. RETRIEVAL OF DEFORMATION FIELDS

The performance of the optimizer for retrieving the deformation fields via transform parameters using the original and the transformed image pairs is illustrated with simulated or designed translation, affine or parameterized B-spline transforms (Fig.7). The original image was transformed using translation, affine and B-spline models as the case may be, and transform was estimated using the registration model (Fig.6, Fig.7 and Fig.8). Exaggerations were introduced in the transform fields that may or may not be common for satellite images but relevant for easy detection and comparisons. The input, retrieved transforms and the difference between them were visualized as raster data, 3D plots, contour plots and histogram data. The translation case is very simple as it is only a constant field and hence seen as a flat field. But affine and B-spline fields vary point to point and the

retrieval of the transform parameters have to be robust which finally depend upon the B-spline knot density and intervals. The Table.2 shows the minimum and maximum values in the deformation fields of input, retrieved and difference between them. It can be easily seen that the differences are very low, which indicate that transform parameters have been very well retrieved. The MI based continuous optimization procedure is able to align satellite images with chosen transform models with sub pixel accuracies of the order up to 0.05 pixels as indicated by the mode values in Table.2.

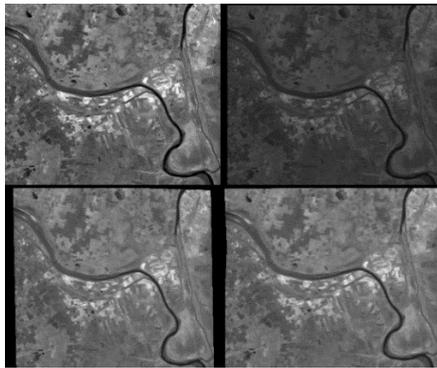


Fig.6. Test image (top left), Translated (top right), Affine (bottom left) and B-spline (bottom right) transformed images

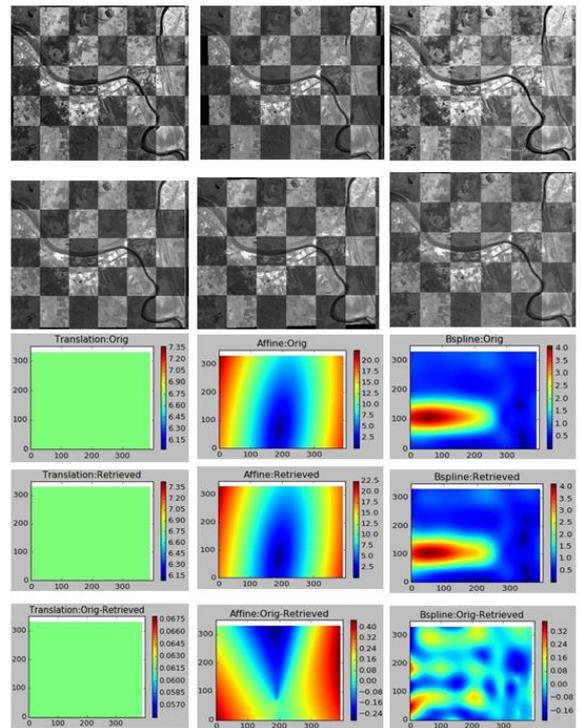


Fig.8. Checker board display before, after registration and transform fields before, after, and difference retrieved

The Fig.8 shows the checker board display before (first row) and after (second row) registration using test image, translation transformed (first column), affine transformed (second column) and B-spline transformed (third column). Third and fourth rows show the deformation fields retrieved and difference between input and retrieved fields for translation, affine and B-spline left to right in order.

Table.2. Statistics of input and retrieved deformation fields as performance measure in pixel units

10	Parameters	Input	Retrieved	Difference
Translation	Min	6.701	6.763	-0.061
	Max	6.701	6.763	-0.061
	Mean	6.701	6.763	-0.061
	Mode	6.701	6.763	-0.061
	STD	$7.8e^{-6}$	$8.4e^{-6}$	0
Affine	Min	0.023	0.021	-0.451
	Max	22.39	22.514	0.285
	Mean	10.39	10.510	-0.120
	Mode	7.583	7.446	-0.049
	STD	4.989	5.101	0.181
B-spline	Min	0.004	0.004	-0.385
	Max	4.107	4.152	0.216
	Mean	1.101	1.103	-0.001
	Mode	0.653	0.596	0.008
	STD	0.917	0.936	0.081

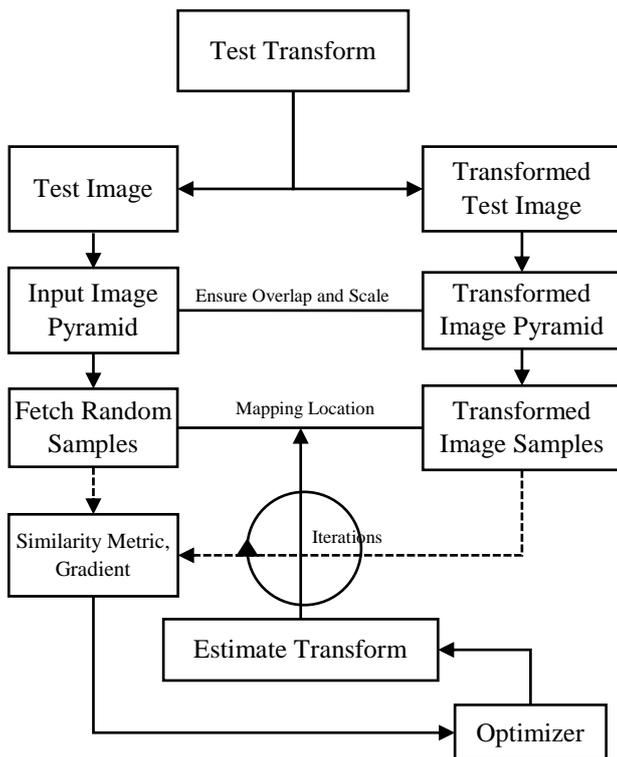


Fig.7. Scheme for retrieval of transform for registration deform field

6. CONCLUSION

Image registration problem formulated as minimization of similarity metric function w.r.t chosen transform model has been proven in recent times as an effective alternative to the long time prevailed, feature matching and estimate transform thereafter employing a least square procedure. Advanced numerical implementation of optimization techniques in which MI serves both as a similarity measure and objective function to be minimized. MI does not require any preprocessing before used in image matching and it provides a sharp measure for finding optimal transformations. Ten percent of samples are enough to estimate cost function and its derivatives to set up optimization loops. Sub pixel accuracies of order 0.05 can be achieved with sufficient number of iterations. Different transform fields such as translation, affine and free form deformations modeled with B-spline can be retrieved reliably using this methodology to a fraction of a pixel unit.

ACKNOWLEDGMENTS

The authors thank Shri. Tapan Misra, Director, Space Applications Centre, ISRO, Ahmedabad who supports us in every scientific endeavor objectively, and Shri. Debjyoti Dhar, Group Director, SIPG, SAC, for his encouragement. We thank SRM Institute of Science and Technology for academic support in research activities.

REFERENCES

- [1] A. Collignon, F. Maes, D. Vandermeulen, P. Suetens and G. Marchal, "Automated Multimodality Medical Image Registration using Information Theory", *Proceedings of 15th International Conference on Information Processing in Medical Imaging*, pp. 263-274, 1995.
- [2] L.G. Brown, "A Survey of Image Registration Techniques", *ACM Computing Surveys*, Vol. 24, No. 4, pp. 325-376, 1992.
- [3] B. Zitova and J. Flusser, "Image Registration Methods: A Survey", *Image and Vision Computing*, Vol. 21, pp. 977-1000, 2003.
- [4] X. Zhen and Z. Yun, "A Critical Review of Image Registration Methods", *International Journal of Image and Data Fusion*, Vol. 1, No. 2, pp. 137-158, 2010.
- [5] T.M. Cover and J.A. Thomas, "Entropy, Relative Entropy and Mutual Information", *Elements of Information Theory*, Wiley and Sons, 1991.
- [6] P. Viola and W.M. Wells, "Alignment by Maximization of Mutual Information", *International Journal of Computer Vision*, Vol. 24, No. 2, pp. 137-154, 1997.
- [7] J.P. Pluim, J.B.A. Maintz and M.A. Viergever, "Mutual-Information-Based Registration of Medical Images: A Survey", *IEEE Transactions on Medical Imaging*, Vol. 22, No. 8, pp. 986-1004, 2003.
- [8] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal and P. Suetens, "Multimodality Image Registration by Maximization of Mutual Information", *IEEE Transactions on Medical Imaging*, Vol. 16, No. 2, pp. 187-198, 1997.
- [9] J. Le Moigne et al., "Multiple Sensor Image Registration, Image Fusion and Dimension Reduction of Earth Science Imagery", *Proceedings of 5th International Conference on Information Fusion*, pp. 999-1006, 2002.
- [10] A.A. Cole-Rhodes, K.L. Johnson, J. LeMoigne and I. Zavorin, "Multiresolution Registration of Remote Sensing Imagery by Optimization of Mutual Information using a Stochastic Gradient", *IEEE Transactions on Image Processing*, Vol. 12, No. 12, pp. 1495-1511, 2003.
- [11] Hua-Mei Chen, R.K. Varshney and M.K. Arora, "Performance of Mutual Information Similarity Measure for Registration of Multitemporal Remote Sensing Images", *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 41, No. 11, pp. 2445-2454, 2003.
- [12] P. Thevenaz and M. Unser, "A Pyramid Approach to Sub-Pixel Image Fusion based on Mutual Information", *Proceedings of IEEE International Conference on Image Processing*, pp. 265-268, 1996.
- [13] M. Unser, "Splines: A Perfect Fit for Signal and Image Processing", *IEEE Signal Processing Magazine*, Vol. 16, No. 6, pp. 22-38, 1999.
- [14] D. Rueckert, L.I. Sonoda, C. Hayes, D.L.G. Hill, M.O. Leach and D.J. Hawkes, "Nonrigid Registration using Free-Form Deformations: Application to Breast MR Images", *IEEE Transactions on Medical Imaging*, Vol. 18, No. 8, pp. 712-721, 1999.
- [15] S. Klein, M. Staring, and J.P.W. Pluim, "Evaluation of Optimization Methods for Non Rigid Medical Image Registration using Mutual Information and B-Splines", *IEEE Transactions on Image Processing*, Vol. 16, No. 12, pp. 2879-2890, 2007.
- [16] S. Klein, J.P.W. Pluim, M. Staring and M.A. Viergever, "Adaptive Stochastic Gradient Descent Optimisation for Image Registration", *International Journal of Computer Vision*, Vol. 81, pp. 227-239, 2009.
- [17] S. Klein, M. Staring, K. Murphy, M.A. Viergever and J.P.W. Pluim, "Elastix: A Toolbox for Intensity-based Medical Image Registration", *IEEE Transactions on Medical Imaging*, Vol. 29, No. 1, pp. 196-205, 2010.
- [18] H.J. Kushner, and G.G. Yin, "*Stochastic Approximation and Recursive Algorithms and Applications*", Springer, 2003.