

B. Sirisha*, B. Sandhya, Chandra Sekhar Paidimarry and A.S. Chandrasekhara Sastry

A Framework for Image Alignment of TerraSAR-X Images Using Fractional Derivatives and View Synthesis Approach

<https://doi.org/10.1515/jisys-2017-0381>

Received July 30, 2017; previously published online February 21, 2018.

Abstract: Conventional integer order differential operators suffer from poor feature detection accuracy and noise immunity, which leads to image misalignment. A new affine-based fractional order feature detection algorithm is proposed to detect syntactic and semantic structures from the backscattered signal of a TerraSAR-X band stripmap image. To further improve the alignment accuracy, we propose to adapt a view synthesis approach in the standard pipeline of feature-based image alignment. Experiments were performed to test the effectiveness and robustness of the view synthesis approach using a fractional order feature detector. The evaluation results showed that the proposed method achieves high precision and robust alignment of look-angle-varied TerraSAR-X images. The affine features detected using the fractional order operator are more stable and have strong capacity to reduce sturdy speckle noise.

Keywords: Image alignment, feature detection, feature description, affine camera model, view synthesis, fractional derivatives.

1 Introduction

The standard pipeline for feature-based image alignment (FIA) involves feature detection, feature description, feature matching, and transformation estimation using a homography matrix [2, 15]. The standard FIA approach fails mainly due to three reasons: (i) complex and challenging characteristics of TerraSAR-X images, (ii) challenging geometric deformation that can arise between source and target image, and (iii) challenges arising due to inherent speckle noise of TerraSAR-X images. Lepetit and Fua [7] demonstrated that generation of supplementary synthetic views from a single-view image helps in improving the performance of image matching in optical images. Morel and Yu [13] combined a view synthesis approach with SIFT extractor and matching. This framework is called Affine SIFT (ASIFT), which improves the matching performance significantly. Mishkin et al. [12] combined view synthesis with an affine covariant detector (Hessian Affine and MSER) and SIFT matching. The resulting matching approach is called MODS, which outperforms the ASIFT approach in terms of the number of true matches and speed. The incorporating view synthesis algorithm in the standard pipeline of FIA has been proved to be effective in finding more precise and correct feature correspondences, aligning difficult matching cases, and also improves the performance of feature detectors like Hessian Affine [10], MSER [9], and SIFT [8, 11], in terms of addressing the range of deformations between the images. All these aforementioned observations are made for optical images and may not be true for TerraSAR-X images. In this paper, we examine and evaluate the concept of view synthesis, which

*Corresponding author: B. Sirisha, Department of Electronics and Communication Engineering, K L University, Vaddeswaram, Andhra Pradesh, India, e-mail: sirishavamsi@gmail.com

B. Sandhya: Department of Computer Science and Engineering, MVSR Engineering College, Rangareddy, Hyderabad, Telangana, India

Chandra Sekhar Paidimarry: Department of Electronics and Communication Engineering, UCE-Osmania University, Hyderabad, Telangana, India

A.S. Chandrasekhara Sastry: Department of Electronics and Communication Engineering, K L University, Vaddeswaram, Andhra Pradesh, India

is of late being used for optical images for aligning complex TerraSAR-X images. It is observed that the incorporating view synthesis approach in feature-based image registration addresses the challenging geometric deformation that can arise between the source and target images. However, in case of TerraSAR-X images, in addition to geometric deformation, we need to address the problems arising due to sturdy speckle noise. To counter the problem of inherent speckle noise, we have developed a fractional derivative-based feature detector to detect robust and affine invariant feature points.

The contributions of our paper are as follows:

- A new fractional derivative-based affine covariant feature detection algorithm is proposed to detect syntactic and semantic structures from the backscattered signal of TerraSAR-X band stripmap images.
- The adaptive TerraSAR-X image alignment using view synthesis (AIAVS) with an affine covariant fractional order feature detector is proposed.

Employing view synthesis for SAR image alignment is a recent development, and the literature is limited. The rest of the paper is organised in a top-down manner. In Section 2, we introduce the AIAVS. Experimental results and analysis for fractional derivative-based affine covariant detector and the AIAVS framework are presented in Section 3. Finally, Section 4 concludes the paper.

2 AIAVS

AIAVS is used to address the challenging geometric deformation that can arise between source and target images. Adaptive image alignment algorithm generates synthetic views of TerraSAR-X images and then applies the standard image alignment pipeline on an entire set of images until the alignment error is less than the threshold. However, in case of TerraSAR-X images, in addition to geometric deformation, we need to address the problems arising due to sturdy speckle noise. As the integer order feature detectors fail to detect more numbers of feature points, we have developed a fractional derivative-based feature detector to counter this problem. In each iteration, features are detected using the proposed fractional derivative-based affine covariant detector. Synthetic views generated in each iteration are different to the next iteration. The rest of the section describes the stages of the AIAVS algorithm.

- Stage 1: Synthetic view TerraSAR-X image generation.
- Stage 2: Local feature extraction using the proposed fractional derivative-based affine covariant detector.
- Stage 3: Two-way nearest neighbour matching strategy.
- Stage 4: Transformation estimation, error estimation, and loop iteration.

2.1 Stage 1: Synthetic View Image Generation

The affine transformation decomposition matrix A was proposed by J. M. Morel:

$$A = \lambda \begin{bmatrix} \cos\psi & -\sin\psi \\ \sin\psi & \cos\psi \end{bmatrix} \begin{bmatrix} t & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{bmatrix}, A = H_{\lambda} R_1(\Psi) T_t R_2(\phi) \quad (1)$$

where homography matrix H is approximated by affine transformation; scale parameter $\lambda > 0$; R_1 , R_2 parameters represent rotations; and $T_t = t > 1$, where t is absolute tilt; $\phi \in (0, \pi)$ is the longitude w.r.t optical axis; and $\Psi \in (0, 2\pi)$ is the camera rotation parameter along with optical axis. Therefore, tilt, longitude, and scale are the three main parameters involved in generating the views.

2.1.1 Algorithm for Generating Synthetic Views

- The Gaussian scale space pyramid is built with Gaussian $\sigma - \sigma_{\text{base}} < 1$.
- The scale space image obtained is rotated using step $\Delta\phi - \Delta\phi_{\text{base}/t}$.

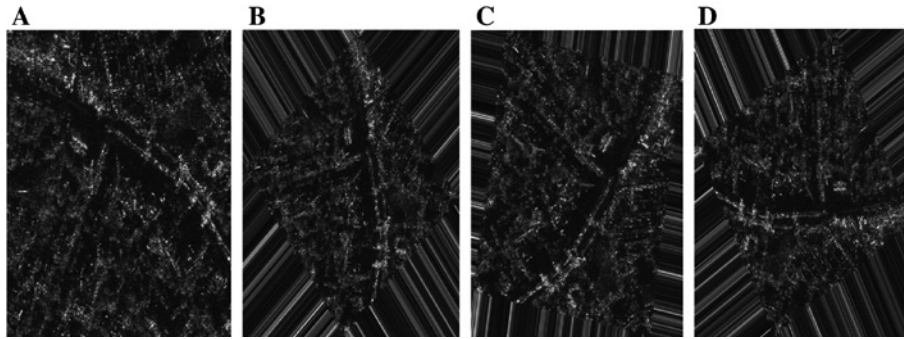


Figure 1: Synthetic Views Generated for Tilt $t = \sqrt{2}$, Angle $\Delta\phi = 72/t = 50^\circ$, $\theta < 180$. Number of views generated is four, i.e. $a = 0^\circ$, $b = 51.4^\circ$, $c = 102.84^\circ$, and $d = 154.2^\circ$.

- The rotated scale space image is convolved with Gaussian filter in horizontal and vertical direction with $(\sigma - \sigma_{\text{base}})$ and $(\sigma - t \cdot \sigma_{\text{base}})$.

Scale λ , latitude ϕ , and longitude θ are the three main parameters that determine the number of views that can be generated in view synthesis. The images first undergo scale change and then rotations followed by tilts. The tilt- t operation is accomplished by employing anti-aliasing filter convolution by a Gaussian filter σ_h in horizontal direction and σ_v in vertical direction, and shrunk in the same direction by t . Figure 1 shows the synthetic views generated for scale $\{\lambda\} = 1$, first tilt $t = \sqrt{2}$, angle $\Delta\phi = 72/t = 50^\circ$, and $\theta < 180$. The number of views generated is four, i.e. $a = 0^\circ$, $b = 51.4^\circ$, $c = 102.84^\circ$, and $d = 154.2^\circ$.

2.2 Local Feature Extraction Using Fractional Order Derivatives

In TerraSAR images, the grey level intensities between the two adjacent pixels are highly correlated due to high self-similarity. These fractal structures of SAR images are very difficult to extract when associated with speckle noise. In this paper, we propose a fractional differential-based feature detector to detect complex fractal features and describe it using RootSIFT [1].

2.2.1 Construction of Fractional Differential Masks

Fractional calculus is non-integer order calculus. It is the branch of calculus that generalises the derivative of a function to non-integer order [3]. The fractional order differential is more efficient in the detection of edges and linear features than integer order differential techniques. The fractional order differential technique highlights high-frequency components and preserves all the low-frequency components of the image. The most commonly used fractional differential operators are the Grünwald-Letnikov, Riemann-Liouville, Weyl-Riesz, Erdélyi-Kober, and Caputo operators [6, 14]. The Grünwald-Letnikov derivative of a function f is defined as

$$D^1 f(x) = \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h}. \quad (2)$$

Iterating this operation yields an expression for the n^{th} derivative of a function.

$$D^n f(x) = \lim_{h \rightarrow 0} h^{-n} \sum_{m=0}^n (-1)^m \binom{-n}{m} f(x - mh). \quad (3)$$

The above expression can be generalised to any non-integer number because

- for any natural number, the calculation of the n^{th} derivative is given by an explicit formula;
- the generalisation of the factorial by the gamma function allows;

$$\binom{n}{m} = \frac{(n)(n+1)\dots(n+m-1)}{\Gamma(m+1)}, \quad (4)$$

which is valid for non-integer values. Let $f(t)$ be the one-dimensional signal with $\Omega = [p, q]$. The support domain Ω is further divided by step interval $h=1$ into n parts, $n = \left\lfloor \frac{s-r}{h} \right\rfloor$. The k^{th} order Grünwald-Letnikov is expressed as

$${}_r^G D_s^k = \lim_{h \rightarrow 0} f_h^{(k)}(t) h^{-k} \sum_{q=0}^n \binom{-k}{q} f(s-q+h), \quad (5)$$

where

$$\binom{-k}{q} = \frac{(-k)(-k+1)\dots(-k+q-1)}{\Gamma(q+1)}, \quad (6)$$

Γ = gamma function. The signal $f(t)$ is now expressed as

$$f^k(t) = f(t) + (-k)f(t-1) + \frac{(-k)(-k+1)}{2}f(t-2) + \frac{\Gamma(-k+1)}{\Gamma(n+1)\Gamma(-k+n+1)}f(t-n). \quad (7)$$

The two-dimensional fractional derivative mask is obtained by linear filtering [5]. Let the x and y coordinates be $x \in [x_1, x_2]$, $y \in [y_1, y_2]$.

The partial order fractional derivative of $f(x, y)$ is

$$\frac{\partial^k f(x, y)}{\partial x^k} \approx f(x, y) + (-k)f(x-1, y) + \frac{(-k)(-k+1)}{2}f(x-2, y) + \dots, \quad (8)$$

$$\frac{\partial^k f(x, y)}{\partial y^k} \approx f(x, y) + (-k)f(x, y-1) + \frac{(-k)(-k+1)}{2}f(x, y-2) + \dots. \quad (9)$$

Using Eqs. (8) and (9), a 3×3 fractional derivative mask is constructed. Figure 2 shows the images obtained after convolving with a fractional mask of order $k=0.1-0.8$. It can be observed from Figure 2 that the characteristics of the image change as the order of k varies. As the order of k increases, the high-frequency responses are highlighted without smoothing the low-frequency components. Hence, the textural, structural components are preserved.

2.3 Proposed Fractional Affine Detector

2.3.1 Algorithm

- Spatial localisation: First-order derivative filters, like Robert, Prewitt, and Sobel, are convolved with fractional order differential filters to obtain a second-order fractional mask. The gradient image of the first-order derivative Sobel filter is given as

$$g_x = -g(x-1, y-1) + g(x+1, y-1) - 2g(x, y-1) + 2g(x, y+1) - g(x-1, y+1) + g(x+1, y+1), \quad (10)$$

$$g_y = -g(x-1, y-1) + g(x-1, y+1) - 2g(x, y) + 2g(x, y+1) - g(x+1, y-1) + g(x+1, y+1). \quad (11)$$

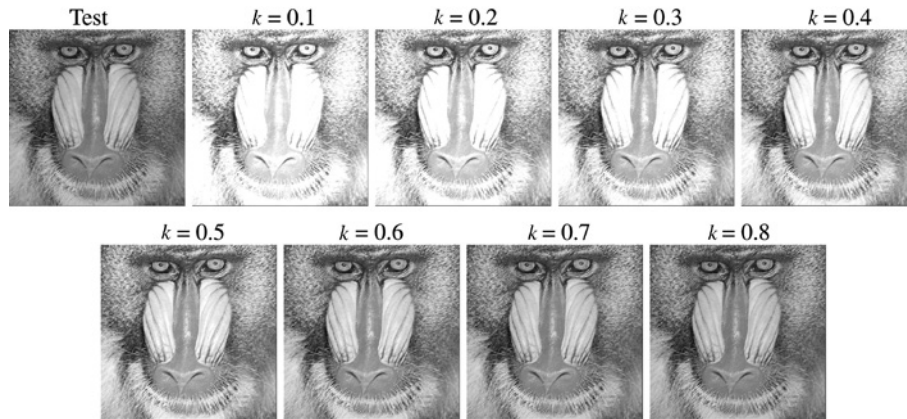


Figure 2: Images Obtained after Convolving the Test Image with the Fractional Mask of Order k (0.1–0.8).

The fractional order differential gradient equations in x and y directions are given below:

$$f(x) = \frac{k^2 - k}{2} f(x-1, y) - kf(x, y) + f(x+1, y), \quad (12)$$

$$f(y) = \frac{k^2 - k}{2} f(x, y-1) - kf(x, y) + f(x, y+1). \quad (13)$$

The gradient image of the first-order derivative Sobel filter g_x in Eq. (10) is convolved with the fractional order differential filter f_x in Eq. (12) to obtain a second-order fractional mask L_x .

The gradient image of the first-order derivative Sobel filter g_y in Eq. (11) is convolved with the fractional order differential filter f_y in Eq. (13) to obtain a second-order fractional mask L_y .

The gradient image of the first-order derivative Sobel filter g_x in Eq. (10) is convolved with the fractional order differential filter f_y in Eq. (13) to obtain a second-order fractional mask L_{xy} .

The gradient image of the first-order derivative Sobel filter g_y in Eq. (11) is convolved with the fractional order differential filter f_x in Eq. (12) to obtain a second-order fractional mask L_{yx} .

Scale-adapted fractional derivative detector:

$$\mu(x, \sigma I, \sigma D) = \begin{bmatrix} \mu_{11} & \mu_{12} \\ \mu_{21} & \mu_{22} \end{bmatrix}, \quad (14)$$

$$= \sigma_D^2 g(\sigma_I) * \begin{bmatrix} L_x L_x(x, \sigma_D) & L_y L_x(x, \sigma_D) \\ L_x L_y(x, \sigma_D) & L_y L_y(x, \sigma_D) \end{bmatrix}. \quad (15)$$

L_x, L_y are fractional second-order image derivatives computed in their corresponding directions using the Gaussian scale σ_D .

The σ_I decides the present scale at which fractional points are detected in Gaussian scale space and also performs weighted averaging of derivatives in eight neighbourhoods. The σ_D -derivative scale decides the Gaussian kernel size.

Cornerness is calculated using the determinant trace of scale-adapted second moment matrix.

Key points are obtained by detecting the local maxima of a point in its eight neighbourhoods. The threshold value is used to filter the poor cornerness points. Figure 3 shows the cornerness image for order $k=0.8$.

- (ii) Each detected feature point is normalised to be affine invariant using affine shape adaptation.
- (iii) The affine region is assessed iteratively by carefully selecting the integration scale, differentiation scale, and spatially localised feature points.

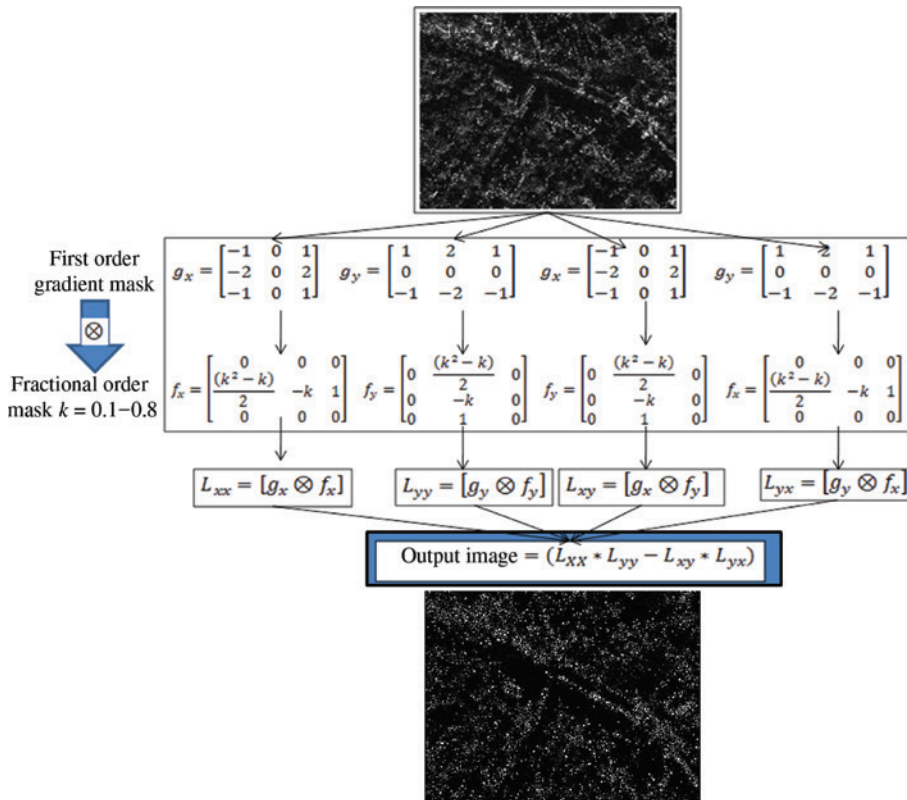


Figure 3: Cornerness Image for ($k=0.8$).

- (iv) The affine region is updated with the selected spatial locations and scales.
- (v) Step (iii) is rerun if the stopping criterion is not met. The stopping criterion is decided based on the eigenvalues of the affine transformation matrix.

It is observed that the fractional differential filter not only maintains the low-frequency contour information in the smooth area, but also highlights the high-frequency edge and texture part in the image. This property has special advantage and visual effect for the images whose texture information has important meaning. Hence, fractional derivative-based affine detector responds very well to textural, structural scenes compared to integer order Hessian Affine detector.

2.4 Stage 3: Two-Way Nearest Neighbour Matching Strategy

The standard first- to second-nearest neighbour ratio matching fails when multiple interpretation of the same features are present. The drawback of multiple detections of the same features is enlarged in case of view synthesis, as Hessian Affine-detected local feature points often have a response in several synthetic views. In order to address this, we use two-way nearest-neighbour matching of the feature descriptors using Bhattacharyya distance.

Let S_x and T_y be the input image pair with corresponding sets of feature points $X=X_i$ and $Y=Y_j$. A match between the pair of feature points is established only if X_i is the perfect match for Y_j in association with all the other feature points in X , and Y_j is the perfect match for X_i in association with all the other feature points in Y .

This matching approach improves the alignment accuracy, as the number of true matches obtained using the two-way matching approach is higher than that obtained using traditional matching. Figure 4 shows the corresponding feature points of two SAR images that vary by a scale factor of 4.5. Figure 4A shows the correspondences obtained using the standard first- to second-nearest neighbour ratio matching strategy, and

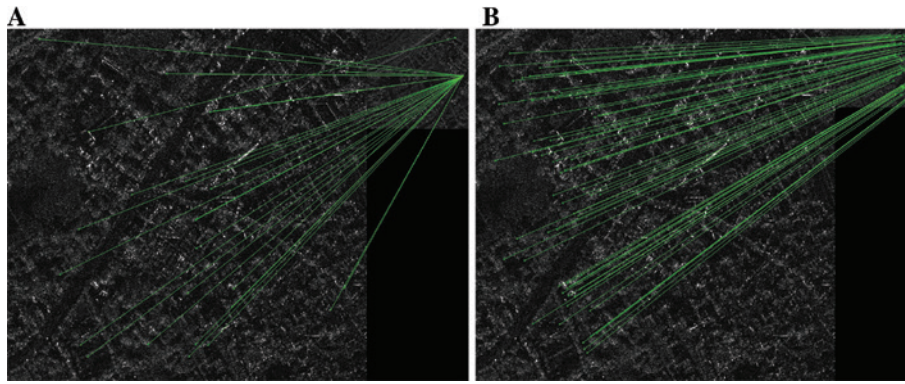


Figure 4: Comparison of the Proposed Two-Way Nearest-Neighbour Ratio Matching Strategy Using (A) Bhattacharya Distance and the (B) Standard First to Second NNR Matching Strategy.

Figure 4B shows the correspondences obtained using the two-way nearest-neighbour ratio (NNR) matching strategy. It is also observed that when the scale factor between the image pair increases, two-way matching is very robust in finding correspondences where traditional approaches fail.

2.5 Stage 4: Geometric Verification, Error Estimation, and Loop Iteration

The main aim of the transformation model is to spatially align the reference and sensed image. To increase the model robustness, outliers should be detected and removed, and only the matched corresponding inlier points should be subjected to transformation. This is achieved using Random Sample Consensus (RANSAC) [4], which detects the inliers of the corresponding feature points and estimates the transformation matrix, H . The source image can be transformed using H to the coordinate system of the target image. Alignment error is computed between the transformed and target images. If the error is above a predefined threshold, the process is iterated with increasing the number of views.

3 Experimental Results and Analysis

The proposed image alignment approach using view synthesis was compared to the standard FIA algorithm and ASIFT algorithm. The experimental results were evaluated and tested on 540 TerraSAR-X images.

3.1 Evaluation Dataset

Though SAR image analysis has been studied extensively, there exists no benchmark dataset of TerraSAR-X band images to compare the performance of various algorithms. Both standard and simulated images are used to evaluate the performance of the proposed prediction model. In this paper, four TerraSAR-X band images, of dimension $10,556 \times 9216$ of the same scene but captured at different look angles, are used for the evaluation. The specification details are as follows: acquisition mode, spotlight 1 m resolution; wavelength, approximately 3 cm; polarisation mode, single; polarizing channel, VV; angle of incidence (look angle), 40.9, 41.9, 42.9, 43.9; date of acquisition, 12 October 2008; look direction, right. Images of size 850×1000 have been cropped from these four images to generate 10 target images. Figure 5 shows 10 target images used in our experiments. From the standard images acquired by the synthetic aperture radar, the dataset is generated by synthetic alteration to incorporate the desired image properties, like scale, rotation, and induced speckle noise.

An area similar to the target image is cropped from the other look-angle SAR images. The source images are generated by applying transformations on the same area cropped from a different look-angle SAR image. One look-angle image is considered the target image, and another look-angle image transformed by applying

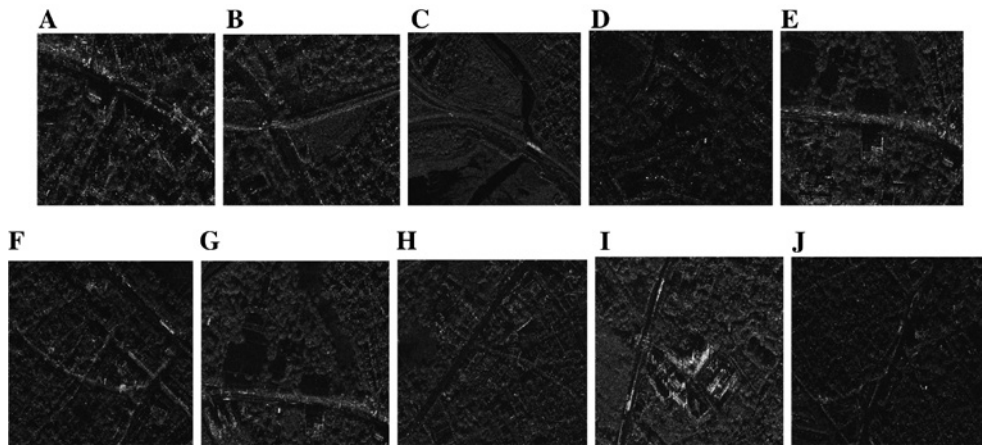


Figure 5: Images Obtained after Convolving with the Fractional Mask of Order k (0.1–0.8) with the Test Image. (A, B) $k=0.1$, (C) $k=0.2$, (D) $k=0.3$, (E) $k=0.4$, (F) $k=0.5$, (G) $k=0.6$, (H) $k=0.7$, and (I) $k=0.8$.

a known geometric transformation is considered the source image. We have created 10 datasets in which the amount of common area (overlap) between the fixed and moving images is varied. The common area between the target and source images in dataset 1, 2 is 100%, in dataset 3 is 95%, in dataset 4 is 90%, in dataset 5 is 85%, in dataset 6 is 80%, in dataset 7 is 75%, in dataset 8 is 70%, in dataset 9 is 60%, and in dataset 10 is 50%. In each dataset, there are 54 SAR images; hence, a total of 540 image pairs are generated, as the pairs differ in varying degrees of scale, rotation, noise, and overlap. The deformation specifications for each dataset are listed in Table 1.

3.1.1 Evaluation of Fractional Affine Feature Detector

The performance of the fractional affine feature detector is tested on dataset 1, which is composed of 54 TerraSAR-X images that vary in look angle, scale, rotation, and speckle noise. The results are analysed in the following ways:

- (i) Selection of suitable first-order derivative filter to be convolved with fractional order derivative filter to obtain a second-order fractional mask.
- (ii) Selection of suitable fractional order- k .

Table 1: Synthesized SAR Image Dataset with Induced Deformation Details.

Type of deformation	Deformation details	Number of image pairs
Look angle	3° Variation	3
Look angle + scale	Scale factor induced between fixed image I_f and moving image I_m Scale factor=0.5; I_f scale down 0.9; I_m scale up 1.8 Scale factor=2; I_f scale up: 1.8; I_m scale down to 0.9 Scale factor=2.5; I_f scale up 1.8; I_m scale down to 0.7 Scale factor=3; I_f scale up 1.8; I_m scale down to 0.6 Scale factor=3.5; I_f scale up 1.8; I_m scale down to 0.51 Scale factor=4.5; I_f scale up 1.8; I_m scale down to 0.4	6
Look angle + rotation	Angle of rotation between I_f and I_m varied from 10° to 350°, with 10° intervals	35
Look angle + speckle noise	Induced speckle noise of variance $v=0.04, 0.05, 0.12, 0.16, 0.2, 0.24, 0.25, 0.32, 0.36, 0.4$	10

Selection of suitable first-order derivative filter

Sobel, Robert, and Prewitt are the three classical first-order derivative filters employed to convolve with fractional order derivative filter. The cornerness image generated using this fractional order detector highlights high-frequency responses and preserves low-frequency components; hence, the quality of the TerraSAR-X image improves by reducing the effect of speckle noise. The performance is measured with the help of peak signal-to-noise ratio (PSNR) and mean square error (MSE). These two measures give the estimate of quality between the original image and the deformed image. Table 2 shows the PSNR and MSE obtained when the Sobel, Robert, and Prewitt operators convolved with the fractional derivative operator for varied orders $k=0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8$, and 0.9 . It is observed from Table 2 that when the Sobel operator convolved with the fractional order derivative $k=0.8$, the value of PSNR is high and the MSE is low compared to the Robert and Prewitt combination. Hence, convolving a Sobel mask with a fractional derivative mask of order $k=0.8$ reduces the effect of speckle noise and gives better feature extraction compared with the fused Robert and Prewitt mask.

Selection of suitable fractional order- k

The fractional order differential operator not only preserves the low-frequency contour information in the smooth area, but also highlights the high-frequency edge and texture part of the images to aid in better feature extraction. However, the optimal selection of the fractional differential order is a crucial problem. Repeatability is a measure used to evaluate feature detector performance, which is used to select the optimal and suitable fractional order. Repeatability is defined as the ratio of the number of matched points to the total number of key points extracted from both images. Figure 6 shows the repeatability values obtained from fractional orders $k=0.2, 0.4, 0.6$, and 0.8 on a pair of TerraSAR-X images deformed by induced speckle noise

Table 2: MSE and PSNR of Prewitt with Fractional Operator (P + F), Robert with Fractional Operator (R + F), and Sobel with Fractional Operator (S + F) for Fractional Orders $k=0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8$, and 0.9 .

k	MSE (P + F)	PSNR (P + F)	MSE (R + F)	PSNR (R + F)	MSE (S + F)	PSNR (S + F)
$k=0.1$	1.4096	8.3931	1.2211	9.1561	1.4779	8.1165
$k=0.2$	1.4326	8.3012	1.216	9.1778	1.4508	8.2213
$k=0.3$	1.3886	8.444	1.2135	9.1883	1.4367	8.2762
$k=0.4$	1.4366	8.2756	1.2226	9.1495	1.4859	8.0858
$k=0.5$	1.5066	7.9947	1.2401	9.0751	1.5691	7.7721
$k=0.6$	1.3924	8.4508	1.2132	9.2898	1.4347	8.2813
$k=0.7$	1.6666	7.3992	1.3176	8.8611	1.7509	7.1167
$k=0.8$	1.3966	8.4623	1.2155	9.3799	1.4481	8.2917
$k=0.9$	2.0766	5.7717	1.5024	8.0228	2.3069	6.3502

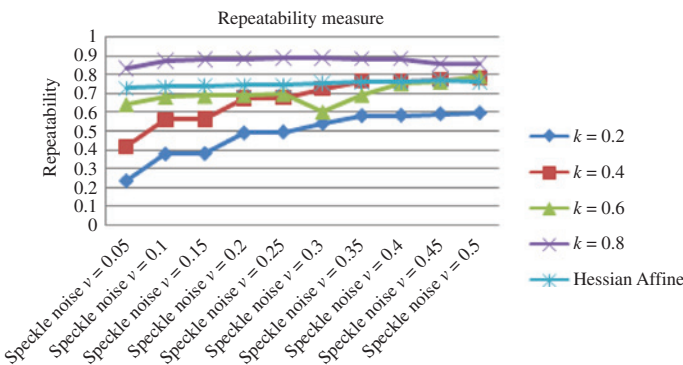


Figure 6: Repeatability Measure for Order $k=0.2, 0.4, 0.6, 0.8$ Compared with Hessian Affine.

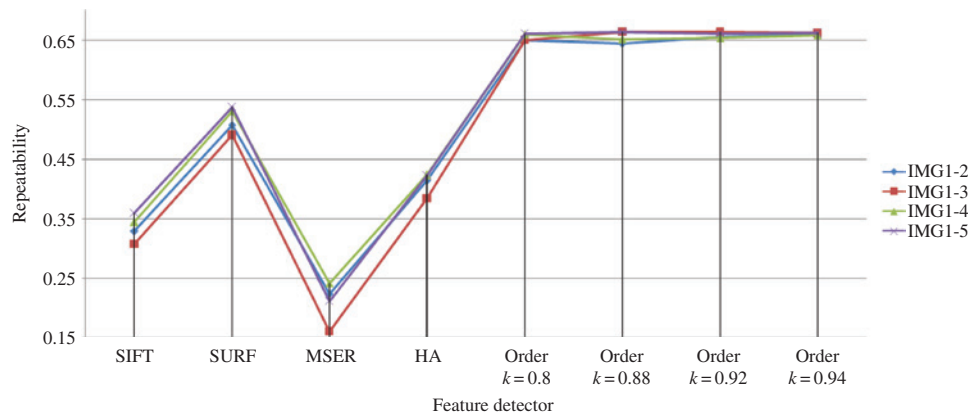


Figure 7: Repeatability Measure for Fractional Order $k=0.8, 0.88, 0.92, 0.94$ is Compared Against Four Integer Order Feature Detectors Like SIFT, SURF, MSER, and Hessian Affine.

of variance 0.05–0.5. The proposed detector performance is compared against the affine invariant integer order detector Hessian Affine. It is observed that for fractional order $k=0.8$, the repeatability value is high and consistent across varied variance values.

We have also tested the performance of the fractional order feature detector on the standard optical image registration dataset Oxford Affine. Five pairs of graffiti images of size 800×640 that vary in view angle with induced speckle noise of variance 0.4 are used for evaluation. A fractional order feature detector of order $k=0.8, 0.88, 0.92$, and 0.94 is compared against four integer order feature detectors like SIFT, SURF, MSER, and Hessian Affine.

It is observed from Figure 7 that for fractional order $k=0.8–0.94$, the repeatability values are almost the same and consistent; hence, fractional order $k=0.8$ is optimal and selected for the experiments to detect features. The affine invariant feature points detected for $k=0.8$ are described using SIFT variant RootSIFT. All the features are stored in a vector array and used for matching.

3.2 Analysis of the AIAVS Approach

The performance of AIAVS is tested on 540 SAR images that vary in look angle, scale, rotation, and speckle noise. The approach is compared against ASIFT and standard FIA. Intel i5 CPU @ 2.6 GHz with 8 Gb RAM, single core machine is used for computations. The results are analysed in the following ways:

- Quantitative assessment is done for 540 synthetically generated SAR images.
- Iteration (tilt) analysis is done deformation wise.
- Qualitative assessment of the AIAVS framework is done.

Table 3 shows the quantitative assessment of the AIAVS framework tested deformation wise. It is observed that out of 540 SAR images, the AIAVS framework could align 502 image pairs, ASIFT could align 320 image pairs, and standard FIA (Hessian Affine detector and SIFT descriptor) could align only 197. It is perceived from Table 3 that the AIAVS framework is effective in addressing extreme rotation and scale deformation images. Figure 8 shows dataset 1 – source SAR image (A); dataset 1–60° rotated target image (B); the number

Table 3: Quantitative Assessment of the AIAVS Approach: Total Number of Aligned SAR Images Deformation Wise.

Total images (540)	Look angle (30)	Rotation (350)	Scale (60)	Speckle (100)	Total aligned images
FIA	30	34	31	100	197
ASIFT	30	150	40	100	320
AIAVS	30	316	56	100	502

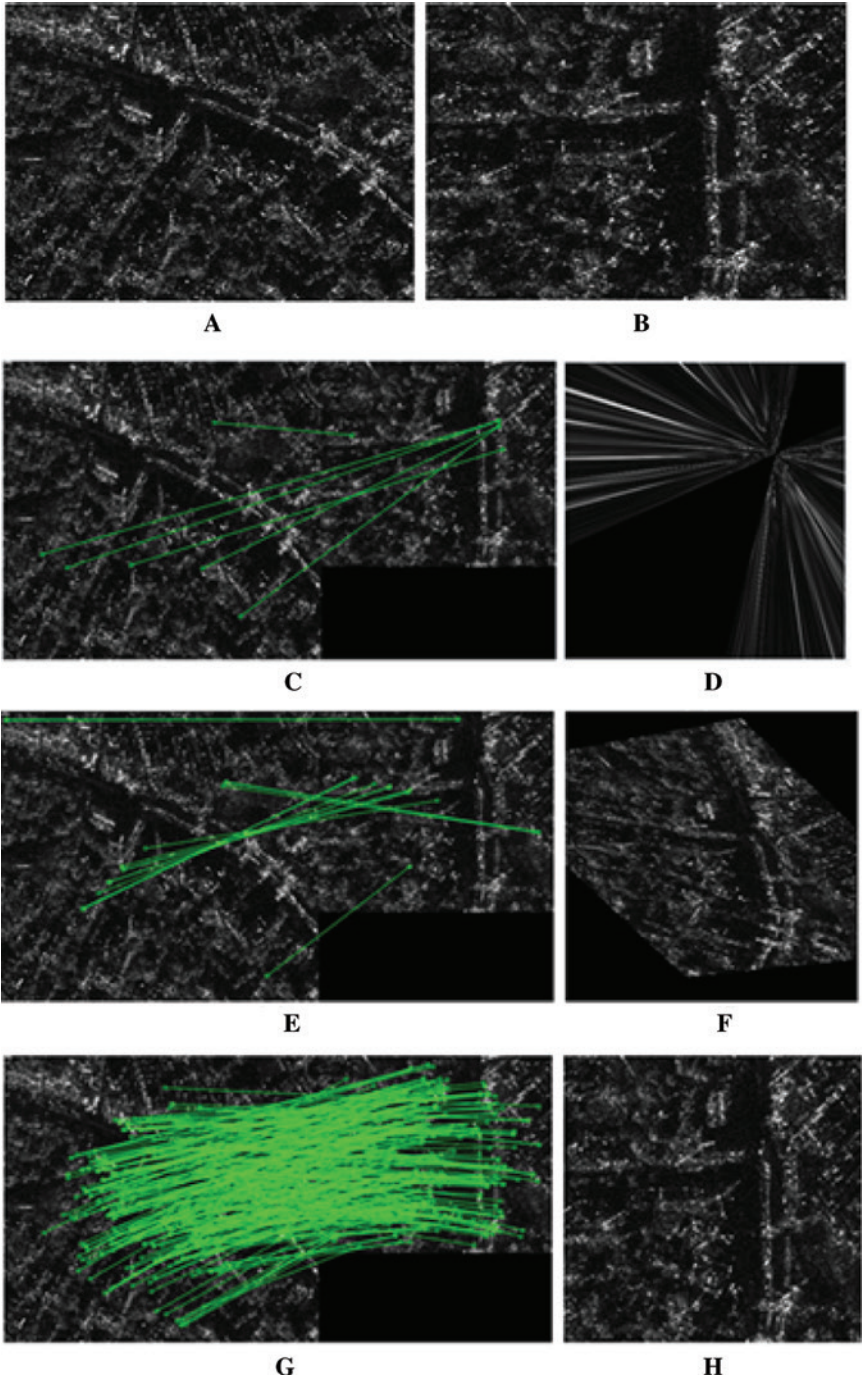


Figure 8: Dataset 1. (A) Dataset 1 – source SAR image; (B) dataset 1 – 60° rotated target image; (C, E, G) number of correspondences obtained using FIA, ASIFT, and AIAVS; and (D, F, H) output aligned images using FIA, ASIFT, and AIAVS.

of correspondences obtained using FIA, ASIFT, and AIAVS (C, E, G); and the aligned or transformed output image using FIA, ASIFT, and AIAVS (D, F, H). Figure 9 shows dataset 2 – source SAR image (A); dataset 2 – scale deformed target image (B); the number of correspondences obtained using FIA, ASIFT, and AIAVS (C, E, G); and the output aligned images using FIA, ASIFT, and AIAVS (D, F, H). It is observed from Figures 8 and 9 that the AIAVS approach can align TerraSAR image pairs varying by any affine deformation.

Table 4 shows the iteration/tilt wise analysis. It is observed that the look angle and induced speckle noise deformed SAR images could be aligned in first tilt. In the case of scale deformation, most of the images below

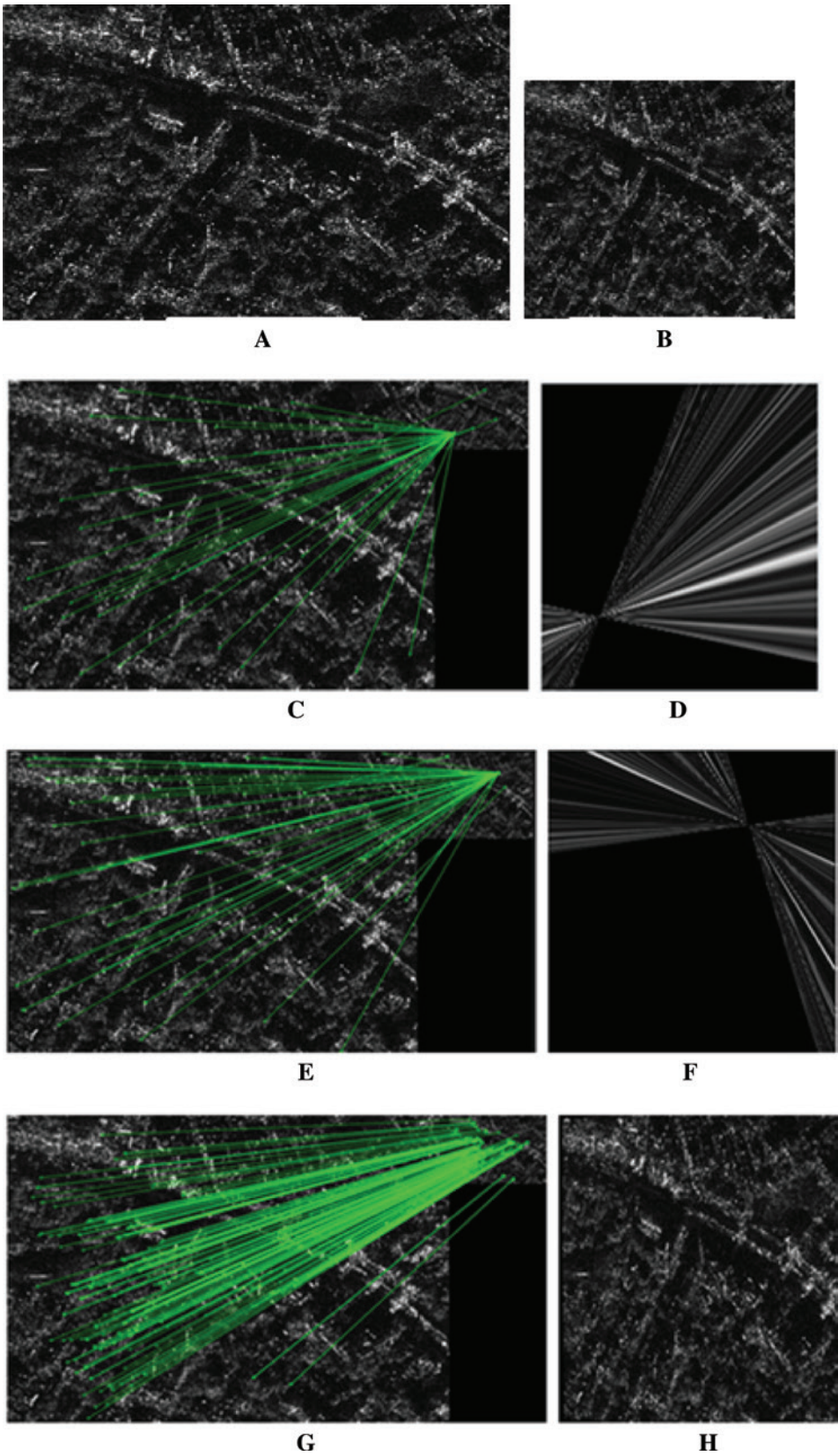


Figure 9: Dataset 2.
(A) Dataset 2 – source SAR image; (B) dataset 2 – scale deformed target image; (C, E, G) number of correspondences obtained using FIA, ASIFT, and AIAVS; and (D, F, H) output aligned images using FIA, ASIFT, and AIAVS.

Table 4: Iteration (Tilt) Analysis: Total Number of Aligned SAR Images Deformation and Iteration/Tilt Wise.

Deformation	Iter-1	Iter-2	Iter-3	Iter-4	Iter-5	Iter-6	Not aligned
Look angle (30)	30	0	0	0	0	0	0
Rotation (350)	179	70	41	12	10	4	32
Scale (60)	37	13	3	3	0	0	4
Speckle (100)	100	0	0	0	0	0	0
Total aligned	346	83	44	15	10	4	36

Table 5: Qualitative Assessment of Number of Inliers (I), Key Point Error (KPE), and Time (Seconds) among the FIA, ASIFT, and Proposed AIAVS Approaches.

	I-FIA	I-ASIFT	I-AIAVS	KPE-FIA	KPE-ASIFT	KPE-AIAVS	Time-FIA	Time-ASIFT	Time-AIAVS
Look angle deformation									
1.2°	280	708	1213	5.98	4.44	0.88	25.8	192	45
1.4°	282	649	1010	6.10	4.65	0.918	26	180	32.5
1.6°	795	2821	3295	2.66	2.24	0.35	38.2	200	52
Rotation deformation									
60°	6	22	382	13.32	6.43	4.089	5.5	120	60
130°	8	20	352	10.23	6.95	3.963	8	132	72
240°	7	15	385	14.11	7.258	4.625	6.9	125	69
Scale deformation									
2	299	27	1176	5.81	9.203	3.27	30	138	54
3	46	68	899	3.88	1.44	2.30	12.3	131	60
4.5	35	206	200	2.90	1.246	1.65	11.8	250	72
Speckle deformation									
$v=0.2$	266	389	1052	6.86	4.69	0.56	24.2	160	43
$v=0.32$	263	547	1100	7.27	4.78	0.59	24	195	38
$v=0.4$	234	573	1142	6.922	4.62	0.63	21.9	189	54

scale factor 2.5 could be aligned in first iteration. When the scale factor increases above 2.5, images become aligned in subsequent iterations. In case of rotation deformation, we observed that for angles between 10–120° and 250–355°, varied images could align in first tilt. In rotation deformation for angles between 120° and 245°, images become aligned in higher iterations. Qualitative assessment of the AIAVS approach is done in terms of number of inliers, key point error, and time (seconds). It is observed from Table 5 that the number of inliers obtained for the AIAVS approach is high and the key point error is low compared to the ASIFT and FIA approaches. The ASIFT algorithm generates a lower number of correct inliers and is slower than AIAVS because it employs standard NNR matching criteria, which eliminate one to many correspondences, including true correspondences.

4 Conclusion

AIAVS is used to address challenging geometric deformation that can arise between source and target images. However, in case of TerraSAR-X images, in addition to geometric deformation, we need to address the problems arising due to sturdy speckle noise. It is observed that integer order feature detectors fail to detect more numbers of feature points; hence, we have developed a fractional derivative-based feature detector to counter this problem. Incorporating fractional-based affine detector in view synthesis approach improves the accuracy of TerraSAR-X image alignment even in the presence of speckle noise.

Bibliography

- [1] R. Arandjelovic, Three things everyone should know to improve object retrieval, in: *Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, CVPR '12, pp. 2911–2918, IEEE Computer Society, Washington, DC, USA, 2012.
- [2] L. G. Brown, A survey of image registration techniques, *ACM Comput. Surv.* **24** (1992), 325–376.
- [3] K. Diethelm and N. J. Ford, *Analysis of Fractional Differential Equations*, Springer, Berlin, Germany, 1999.
- [4] M. A. Fischler and R. C. Bolles, Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography, *Commun. ACM* **24** (1981), 381–395.
- [5] S. Kempfle, I. Schäfer and H. Beyer, Fractional calculus via functional calculus: theory and applications, *Nonl. Dynam.* **29** (2002), 99–127.
- [6] B. T. Krishna, Studies on fractional order differentiators and integrators: a survey, *Signal Process.* **91** (2011), 386–426.
- [7] V. Lepetit and P. Fua, Keypoint recognition using randomized trees, *IEEE Trans. Pattern Anal. Mach. Intell.* **28** (2006), 1465–1479.
- [8] D. G. Lowe, Distinctive image features from scale-invariant keypoints, *Int. J. Comput. Vision* **60** (2004), 91–110.
- [9] J. Matas, O. Chum, M. Urban and T. Pajdla, Robust wide baseline stereo from maximally stable extremal regions, in: P. L. Rosin and A. D. Marshall, eds., *BMVC*, British Machine Vision Association, Durham, UK, 2002.
- [10] K. Mikolajczyk and C. Schmid, Scale & affine invariant interest point detectors, *Int. J. Comput. Vision* **60** (2004), 63–86.
- [11] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir and L. Van Gool, A comparison of affine region detectors, *Int. J. Comput. Vision* **65** (2005), 43–72.
- [12] D. Mishkin, J. Matas and M. Perdoch, MODS: fast and robust method for two-view matching, *Comput. Vis. Image Understand.* **141** (2015), 81–93.
- [13] J. -M. Morel and G. Yu, ASIFT: a new framework for fully affine invariant image comparison, *SIAM J. Imaging Sci.* **2** (2009), 438–469.
- [14] R. Scherer, S. L. Kalla, Y. Tang and J. Huang, The Grünwald-Letnikov method for fractional differential equations, *Comput. Math. Appl.* **62** (2011), 902–917.
- [15] B. Zitová and J. Flusser, Image registration methods: a survey, *Image Vis. Comput.* **21** (2003), 977–1000.