ROBUST FEATURE MATCHING FOR GEOSPATIAL IMAGES VIA AN AFFINE-INVARIANT COORDINATE SYSTEM

JIAYUAN LI (ljy_whu_2012@whu.edu.cn)
QINGWU HU* (huqw@whu.edu.cn)
MINGYAO AI (aimingyao@whu.edu.cn)

School of Remote Sensing and Information Engineering, Wuhan University, China

*Corresponding author

Abstract

Feature matching is a crucial stage for many photogrammetric and remote sensing applications. This paper proposes a novel and robust feature-matching method based on a normalised barycentric coordinate system (NBCS), which is superior to a Cartesian system for this task. A scale-invariant feature transform (SIFT) is performed to provide initial matches containing both correct matches (inliers) and false matches (outliers), with a focus on model estimation from contaminated observations (matches with outliers). An affine-invariant coordinate system called NBCS is defined based on ratios of areas. The two feature points of a correct match have the same coordinates under NBCS while false correspondences do not. This principle is adapted into a hypothesise-and-verify framework. The proposed method is robust, efficient and effective. Extensive experiments on real geospatial image pairs show that it significantly outperforms six other state-of-the-art approaches. The source code and datasets used in this paper have been made public.[1]

KEYWORDS: affine, hypothesise-and-verify, image registration, normalised barycentric coordinate system (NBCS), outlier filtering, robust feature matching

INTRODUCTION

ROBUST FEATURE MATCHING, used to establish reliable correspondences in overlapping image pairs, is a crucial step for many photogrammetric and remote sensing applications, such as image registration, image orientation and structure from motion. Robust feature-matching methods usually consist of two major stages, namely, initial feature matching and outlier elimination (Ma et al., 2015). In the method proposed in this paper, the focus is on improving the outlier filtering step.

Feature-matching methods detect distinct structures, including feature points, lines and regions, and usually seek correspondences between local features by computing their

---

[1] https://sites.google.com/site/jiayuanli2016whu/home

descriptor vector distances. The scale-invariant feature transform (SIFT) (Lowe, 2004) is the most well-known method in this category. Many variants of SIFT have been developed for geospatial image matching and registration. For instance, Sedaghat et al. (2011) proposed a variant called *uniform robust SIFT* by a novel feature-selection strategy to improve feature distribution. Dellinger et al. (2015) proposed a method called SAR-SIFT for synthetic aperture radar (SAR) image matching. It presented a new gradient definition which is robust to speckle noise. Similarly, Ma et al. (2017) also introduced a new gradient definition and proposed an enhanced feature-matching method called PSO-SIFT by combining the position, scale and orientation of each keypoint. These methods can establish distinct initial feature correspondences for image registration. However, outliers are still preserved because of the illumination, viewpoint, rotation, scale, temporal or speckle noise differences between image pairs.

The second step is then adopted to clean the initial feature correspondences. Outlier elimination approaches can be roughly classified into two categories, namely, parametric-based methods and non-parametric ones (Ma et al., 2014; 2015; Li et al., 2016).

Parametric-based methods usually use a transformation model, such as a conformal, affine or projective (homography) transformation, to describe the geometric relationship between an image pair, and to estimate such a model from feature correspondences contaminated by outliers. Random sample consensus (RANSAC) (Fischler and Bolles, 1981) and its variants (Torr and Zisserman, 2000; Chum et al., 2003; Chum and Matas, 2005; Lebeda et al., 2012; Moisan et al., 2012; Hast et al., 2013; Raguram et al., 2013; Wu et al., 2015) use a hypothesise-and-verify technique for this problem. They alternate between minimum subset selection and transformation model verification. The model with the most supporting correspondences is accepted as the correct solution. For example, Torr and Zisserman (2000) presented maximum likelihood estimation sample consensus (MLESAC) as a generalised robust estimator of RANSAC, which maximised the likelihood based on probability. Chum and Matas (2005) used local similarity ordering in their progressive sample consensus (PROSAC) algorithm to draw the minimal subset of correspondences, which improved the first stage of RANSAC. Locally optimised RANSAC (LO-RANSAC) (Chum et al., 2003) introduced a local optimisation stage to improve RANSAC, which significantly decreased the number of samples drawn. Lebeda et al. (2012) introduced a truncated quadratic cost function in the local optimisation procedure of LO-RANSAC. *A contrario* RANSAC (AC-RANSAC) (Moisan et al., 2012) uses the so-called *a contrario* methodology in order to find a model that best fits the data with a confidence threshold that adapts automatically to noise. Raguram et al. (2013) proposed USAC as a universal framework for RANSAC-like robust feature matching. It extended the basic RANSAC to incorporate a number of important practical and computational considerations. Similarly, optimal RANSAC (Hast et al., 2013) also put several typical RANSAC-like methods together to produce an algorithm that is repeatable. The main limitation of these RANSAC-like methods is that they tend to degrade badly if the outlier ratio of the initial matches becomes large (Li and Hu, 2010; Ma et al., 2014). Many researchers have introduced affine invariants as geometric constraints for image registration. For instance, Li and Ye (2012) introduced triangle-area representation (TAR) to improve the traditional RANSAC algorithm. Zhang et al. (2014) proposed a feature-point descriptor calculated by the TAR of the *k*-nearest neighbours (KNN-TAR). Yang and Cohen (1999) used an affine invariant to develop a class of ordered local affine-invariant features based on the convex hull. More recently, certain direct methods were presented. Locally linear transforming (LLT) by Ma et al. (2015) adapted the Bayesian model to describe the outlier elimination problem and adopted a local geometric constraint within this model. The closed-form

solution was then estimated by expectation maximisation (EM) (Dempster et al., 1977). Li et al. (2016) introduced a new cost function based on the lq-norm for robust feature matching, directly estimating the affine transformation from initial observations contaminated by outliers.

Non-parametric methods are widely used in computer vision applications. Generally, they are suitable for both rigid and non-rigid image registration. Graph-matching methods use graphs to organise extracted features and minimise the structural distortions between graph networks via an energy function. For instance, Torresani et al. (2008) introduced a technique called dual decomposition for graph matching, in which they defined a complex cost function based on the spatial arrangement, texture similarity and geometric consistency of the keypoints. Cho et al. (2014) introduced a max-pooling strategy to graph matching. In their method, candidate matches were scored by their most promising neighbours, and the scores were then used to update the neighbours. Graph-based methods can be also applied in geospatial image matching, such as Liu et al. (2012) who introduced two-way spatial-order constraints and two decision-criteria restrictions into their graph-matching framework. Hu et al. (2015) used spatial relationships and geometrical constraints for reliable feature matching. However, graph-matching methods usually suffer from high computational complexity. Vector field consensus (VFC) (Ma et al., 2014) formulates the problem as a maximum likelihood estimation of a Bayesian model and estimates a consensus of inliers based on a vector field.

In this paper, a robust parametric-based method is proposed for geospatial image matching and registration. What is different from the above-mentioned affine-invariant methods is that a coordinate system is derived, called the normalised barycentric coordinate system (NBCS), based on ratios of areas. Any image points with Cartesian coordinates can be transferred into normalised barycentric coordinates (NBC), which are chosen because they are superior to a Cartesian coordinate system for this task. NBCS is invariant to affine transformation. Thus, two feature points forming a correct match have the same coordinates under NBCS. This principle is adapted into a hypothesise-and-verify framework to improve the robustness and efficiency of traditional methods such as RANSAC. Real experiments on many geospatial image pairs demonstrate that the proposed method is effective, efficient and robust.

ROBUST FEATURE MATCHING

This section describes the proposed robust feature-matching method. A definition of NBCS is presented first and then it is adapted into a hypothesise-and-verify framework for robust transformation estimation.

Normalised Barycentric Coordinate System (NBCS)

Four coplanar points A_1, B_1, C_1, D_1, where any three points are non-collinear (Fig. 1) can form four triangles. The barycentric coordinates $(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$ of the quadrilateral $A_1B_1C_1D_1$ are defined as the ratios of the areas of these triangles:

$$\lambda_1 : \lambda_2 : \lambda_3 : \lambda_4 = S_{\Delta A_1 B_1 C_1} : S_{\Delta A_1 B_1 D_1} : S_{\Delta A_1 C_1 D_1} : S_{\Delta B_1 C_1 D_1} \tag{1}$$

where $S_{\Delta A_i B_i C_i}$ $(i = 1, 2, 3, 4)$ represents the area of triangle $\Delta A_i B_i C_i$ $(i = 1, 2, 3, 4)$. The barycentric coordinates are then normalised to (u_1, u_2, u_3, u_4), where $u_i = \lambda_i/s$ $(i = 1, 2, 3, 4)$ are the NBCS coordinates and $s = \lambda_1 + \lambda_2 + \lambda_3 + \lambda_4$.
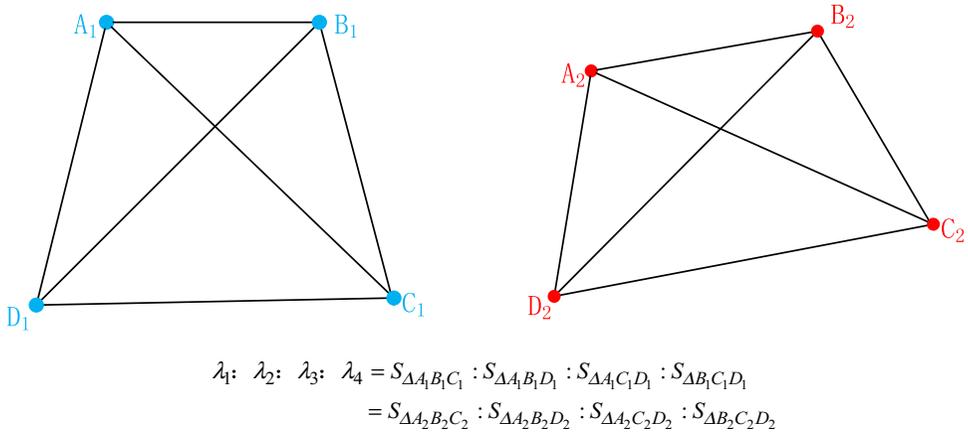
$$\lambda_1: \ \lambda_2: \ \lambda_3: \ \lambda_4 = S_{\Delta A_1 B_1 C_1} : S_{\Delta A_1 B_1 D_1} : S_{\Delta A_1 C_1 D_1} : S_{\Delta B_1 C_1 D_1}$$
$$= S_{\Delta A_2 B_2 C_2} : S_{\Delta A_2 B_2 D_2} : S_{\Delta A_2 C_2 D_2} : S_{\Delta B_2 C_2 D_2}$$

Fɪɢ. 1. The invariance of barycentric coordinates. $A_1 B_1 C_1 D_1$ is a quadrilateral and $A_2 B_2 C_2 D_2$ is the corresponding quadrilateral after applying an affine transformation T. Four triangles can be formed from each quadrilateral and its barycentric coordinates are the area ratios of these triangles. The normalised barycentric coordinates (NBC) of the correspondence are equal.

As pointed out by Hartley and Zisserman (2003), an affine matrix $\mathbf{A}$ consists of two terms, that is, the rotation term $\mathbf{R}(\theta)$ and the deformation term $\mathbf{R}(-\phi)\mathbf{D}\mathbf{R}(\phi)$:

$$\mathbf{A} = \mathbf{R}(\theta)\mathbf{R}(-\phi)\mathbf{D}\mathbf{R}(\phi) \tag{2}$$

where $\mathbf{D}$ is formed by scale parameters σ_1 and σ_2,

$$\mathbf{D} = \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix}. \tag{3}$$

An affine transformation includes translations, rotations and non-isotropic scaling. An area is invariant to translations and rotations, so it is only affected by the scaling σ_1 and σ_2. The area is scaled by $\sigma_1 \cdot \sigma_2$ which is equal to $\det(\mathbf{A})$. The ratio of areas will eliminate the scale term $\det(\mathbf{A})$ and be invariant under affine transformations. Thus, the NBC of each quadrilateral is invariant to affine transformations and is unique.

Fig. 1 indicates that points A_1, B_1, C_1, D_1 are transformed to points A_2, B_2, C_2, D_2 by an affine transformation T. Based on its definition, the NBC (u'_1, u'_2, u'_3, u'_4) of the quadrilateral $A_2 B_2 C_2 D_2$ is:

$$(u'_1, u'_2, u'_3, u'_4) = \frac{1}{s'} \cdot (\lambda'_1, \lambda'_2, \lambda'_3, \lambda'_4) = \frac{\det(\mathbf{A})}{s'} \cdot (\lambda_1, \lambda_2, \lambda_3, \lambda_4)$$
$$= \frac{\det(\mathbf{A})}{\det(\mathbf{A}) \cdot s} \cdot (\lambda_1, \lambda_2, \lambda_3, \lambda_4) = (u_1, u_2, u_3, u_4) \tag{4}$$

where $(\lambda'_1, \lambda'_2, \lambda'_3, \lambda'_4)$ are the barycentric coordinates of the quadrilateral $A_2 B_2 C_2 D_2$, and $s' = (\lambda'_1 + \lambda'_2 + \lambda'_3 + \lambda'_4) = \det(\mathbf{A}) \cdot s$.

Once a correct quadrilateral correspondence $(A_1 B_1 C_1 D_1, A_2 B_2 C_2 D_2)$ is established, any image point p_1 in the Cartesian coordinate system in image I_1 can be transferred to NBCS based on $A_1 B_1 C_1 D_1$ (Fig. 2). The NBC of p_1 is:
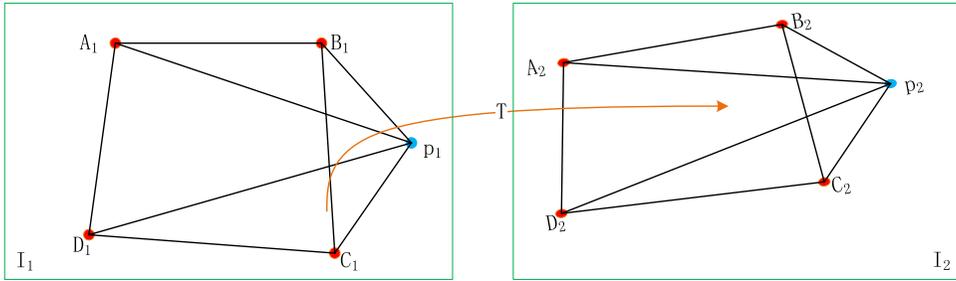
FIG. 2. The normalised barycentric coordinate system (NBCS). Points p_1 in image I_1 and p_2 in image I_2 have the same coordinates in the NBCS under the affine transformation T.

$$\begin{cases} \text{NBC}(p_1) = (S_{p_1A_1B_1}, S_{p_1B_1C_1}, S_{p_1C_1D_1}, S_{p_1A_1D_1})/\rho \\ \rho = (S_{p_1A_1B_1} + S_{p_1B_1C_1} + S_{p_1C_1D_1} + S_{p_1A_1D_1}) \end{cases}. \tag{5}$$

In the same way, any image point p_2 in image I_2 can also be transferred to NBCS based on $A_2B_2C_2D_2$. As NBCS is invariant to affine transformations, the two feature points forming a correct match have the same coordinates under NBCS. The proposed coordinate system is superior to Cartesian coordinates for such a task due to this important property.

Hypothesise-and-Verify with the NBC Principle

Given a geospatial image pair (I_1, I_2), firstly extract N initial correspondences $C = \{(\mathbf{x}_n, \mathbf{y}_n)\}_{n=1}^{N}$ by SIFT, where $\mathbf{x}_n$ and $\mathbf{y}_n$ are the image coordinates of feature points in images I_1 and I_2. The goal of robust feature matching is to establish reliable matches, in other words, distinguishing inliers from outliers. The local surface model described by a geospatial image can be treated approximately as a plane compared with the flying/orbital height of the camera, especially for satellite and high-altitude aerial images (Li et al., 2016). Many researchers use a 2D affine transformation to approximately model the geometric relationship between a geospatial image pair. Thus, $(\mathbf{x}_n, \mathbf{y}_n)$ satisfies the following equation if it is an inlier:

$$\mathbf{y}_n = T(\mathbf{x}_n) = \mathbf{A}\mathbf{x}_n + \mathbf{t} \tag{6}$$

where $\mathbf{A}$ is a 2×2 affine matrix and $\mathbf{t}$ is a 2×1 translation column vector.

As noted earlier, NBC is an invariant under an affine transformation. Thus, if four correct correspondences are selected from a correspondence set C, ideally, the correspondence quadrilateral must satisfy equation (4). In practice, a geospatial image pair does not strictly satisfy an affine transformation and thus the selected correspondence quadrilateral should satisfy the following relationship:

$$\sqrt{\sum_{i=1}^{4}(u_i - u_i')^2} < \delta \tag{7}$$

where δ is a small value that checks if a correspondence has nearly the same NBC. This principle is adapted into a hypothesise-and-verify framework and develops the proposed

robust feature-matching algorithm. First, four feature correspondences are randomly picked and a check made as to whether the invariance of NBC is satisfied. Second, if equation (7) is satisfied, then an affine transformation can be estimated and this solution verified with the remaining matches. The proposed method alternately performs these two stages until a confident solution is determined. The SIFT algorithm usually provides hundreds of initial correspondences with matching scores. However, an affine transformation only has six degrees of freedom. Too many correspondences will not improve the accuracy while decreasing the efficiency. Thus, sample M $(M < N)$ correspondences with the best matching scores from set C to estimate the transformation model T, and then apply T on C to find the inlier set:

$$C_{inlier} = \{(\mathbf{x}_n, \mathbf{y}_n)\}_{n=1}^{K} \text{ s.t. } \mathbf{y}_n - T(\mathbf{x}_n) < \varepsilon$$

where ε is an outlier rejection threshold ($\varepsilon = 3$ in these experiments). The method is summarised in Algorithm 1.

Algorithm 1: Robust feature matching based on barycentric coordinates

---

Input: a geospatial image pair (I_1, I_2)

Output: reliable matches $C_{inlier} = \{(\mathbf{x}_n, \mathbf{y}_n)\}_{n=1}^{K}$,

        transformation T

1 Generate initial matches $C = \{(\mathbf{x}_n, \mathbf{y}_n)\}_{n=1}^{N}$ by SIFT on (I_1, I_2);

2 Sample $M(M \leq N)$ best matches $C_{sampled} \subseteq C$ according to their

  matching scores;

3 **repeat**

4   *Hypothesise*-step:

5   Randomly pick 4 correspondences from $C_{sampled}$;

6   Compute their NBCs;

7   *Verify*-step:

8   **if** equation (7) is true

9   Compute transformation T_i, verify T_i by C

10   **else**

11   Continue;

12   **end**

13 **until** find a confident solution T

14 Apply T on C to find C_{inlier}

EXPERIMENTAL RESULTS

This section reports performance comparisons between the proposed method and the following six other state-of-the-art methods:

(1) LLT – locally linear transforming (Ma et al., 2015);
(2) VFC – vector field consensus (Ma et al., 2014);
(3) RANSAC – random sample consensus (Fischler and Bolles, 1981);
(4) MLESAC – maximum likelihood estimation sample consensus (Torr and Zisserman, 2000);
(5) AC-RANSAC – *a contrario* RANSAC (Moisan et al., 2012); and
(6) FSC – fast sample consensus (Wu et al., 2015).

For LLT, RANSAC, MLESAC and FSC, this research used an affine transformation as the geometric model; for AC-RANSAC, a homography matrix was used; VFC is a non-parametric method. The outlier rejection threshold of RANSAC, MLESAC, FSC and the proposed method was set to 3 pixels. Note that all the values reported in this paper are the average values of 20 image pairs with 50 independent tests (Table I). Parameters for each method are tuned to provide the best result and then fixed. All the experiments were performed on the same laptop (i5, 2·5 GHz Intel Core, 8 GB memory) and all the methods were implemented using MATLAB, except AC-RANSAC (implemented by C++).

Dataset and Setting

To evaluate the proposed method, 20 typical geospatial image pairs were used. The details are summarised in Table I, and thumbnails of these image pairs are shown in Fig. 3. The ground sample distance (GSD) of these images varies from 0·5 to 30 m, including high-, medium- and low-resolution geospatial images. These 20 image pairs suffer from substantial distortions. For instance, local geometric distortions are serious in multitemporal image pairs; the photometric information is significantly different between multisensor image pairs; and the overlapping regions of aerial image pairs 10 to 20 are extremely small.

For each image pair, five evenly distributed matches are manually selected with sub-pixel accuracy to estimate an accurate affine transformation to act as ground truth. VLFeat (an open-source library of computer vision algorithms by Vedaldi and Fulkerson, 2010) is used to perform SIFT for initial correspondence generation. The matching distance threshold for SIFT is set to 1·2 in all experiments. The matches with small residuals (smaller than $\varepsilon = 3$) after ground-truth transformation are chosen as ground-truth inliers. The same initial SIFT matches are used for all seven methods. Precision, recall and f-score are adopted as the evaluation metrics, where *precision* is the percentage of the true inliers in all detected matches and *recall* is computed by dividing the detected inlier number by the ground-truth inlier number. The *f-score* combines both recall and precision metrics into a single metric that reflects the overall performance:

$$f\text{-}score = \frac{2 \cdot precision \cdot recall}{precision + recall} \times 100\%. \tag{8}$$

The maximum error (ME) and root mean square error (RMSE) of the proposed method are also reported.

Parameter and Feature-matching Study

There are two important parameters in the proposed method, namely, the number of sampled matches M and the small tolerance threshold for the NBC constraint δ. To determine the best parameters, firstly, M was fixed at 100 and δ was incremented from 0·01 to 0·1 in steps of 0·01. For each value of δ, the proposed method was performed on the dataset, and the precision, recall, f-score and running time metrics were calculated (Table II). As can be seen, the results only have very small differences when δ changes from 0·01 to 0·05. The best f-score is achieved when $\delta = 0\ 03$. Then, δ was fixed at 0·03 and M changed from 25 to 200 in increments of 25 (Table III). As shown, the f-score increases as M increases until $M = 100$ and then it slightly decreases as M increases further. Generally, more data is better. However, in the proposed method, the correspondences were

TABLE I. Input geospatial image pairs. Pairs 1 to 8 use satellite imagery; pairs 10 to 20 are aerial; pair 9 is composite.

| No.	Image pair platform	Image size (pixels)	GSD (m)	Date	Location
1	WorldView-2	405 × 350	0·5	2011	USA –
	WorldView-2	405 × 350	0·5	2014	California
2	Landsat TM	512 × 512	30	1992	Brazil –
	Landsat TM	512 × 512	30	1994	Amazon
3	JERS-1	256 × 256	18	1995	Brazil –
	JERS-1	256 × 256	18	1996	Amazon
4	Landsat TM	512 × 512	30	1990	USA –
	Landsat TM	512 × 512	30	1994	Iowa
5	SPOT-2	256 × 256	20	1995	Brazil –
	Landsat TM	256 × 256	30	1994	Brasilia
6	SPOT-5	800 × 800	2·5	2002	China –
	SPOT-6	800 × 800	1·5	2012	Beijing
7	SPOT-5	800 × 800	2·5	2003	France –
	SPOT-7	800 × 800	1·5	2014	Paris
8	SPOT-5	1000 × 1000	2·5	2008	China –
	SPOT-5	1000 × 1000	2·5	2012	Shanghai
9	Radarsat-2	800 × 800	3	2013	China –
	Airborne SAR	800 × 800	3	2013	Jiangsu
10–20	Aerial photographs	1391 × 1374 to 1459 × 1380	0·5	2011	USA – Illinois
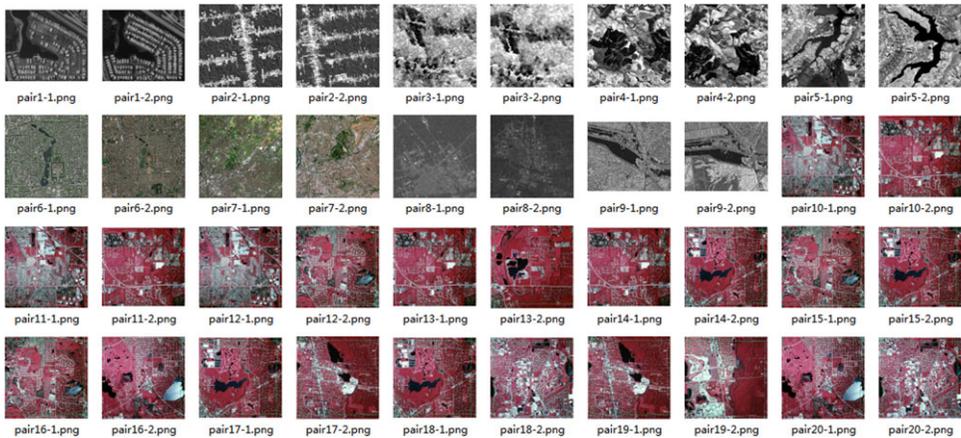
FIG. 3. Thumbnail images of the datasets in Table I.

sampled based on their matching scores. Thus, the sampled correspondences do not have the same precision. Next, $\delta = 0\cdot03$ and $M = 100$ were fixed in the subsequent experiments. Note that $M = 100$ is the best for this dataset. However, if the outlier ratio is rather high,

TABLE II. Study of parameter δ, with *M* fixed at 100. Best overall result is shown in bold.

Metric	δ (*M* = 100)							
	0·01	0·02	**0·03**	0·04	0·05	0·06	0·07	0·08
Precision (%)	97·2	97·4	**97·3**	96·8	96·8	95·5	95·5	94·8
Recall (%)	96·1	96·2	**96·9**	96·0	95·5	93·8	92·8	91·1
f-score (%)	96·1	96·3	**96·9**	96·2	95·9	94·3	93·6	92·2
Time (s)	2·4	1·4	**1·2**	1·0	0·7	0·6	0·5	0·5

TABLE III. Study of parameter *M*, with δ fixed at 0·03. Best overall result is shown in bold.

Metric	*M* (δ = 0·03)							
	25	50	75	**100**	125	150	175	200
Precision (%)	91·3	92·4	96·0	**97·2**	97·0	96·9	96·6	96·4
Recall (%)	88·2	92·7	95·9	**96·5**	95·8	95·7	95·4	94·3
f-score (%)	88·2	92·4	95·7	**96·5**	96·1	96·0	95·7	94·9
Time (s)	5·3	2·7	1·9	**1·3**	1·8	1·9	2·0	2·4

such as 10 000 initial matches with only 100 correct matches, *M* = 100 may not contain a set that passes the NBCS test. Thus, a fine-to-coarse strategy is adopted where, initially, *M* is set to 100. If a good solution can be obtained, the procedure outputs the results; otherwise, the value of *M* is increased by a factor of three and Algorithm 1 is performed again until a good solution is found.

To study the robustness to different initial feature-matching methods, SIFT (Lowe, 2004), PSO-SIFT (Ma et al., 2017) and SAR-SIFT (Dellinger et al., 2015) were used to generate initial feature correspondences (Table IV). As shown, different initial feature-matching methods have almost no influence on the proposed method.

NBCS is the major contribution of this paper. To prove that NBCS outperforms the Euclidean coordinate system for the geospatial robust feature-matching problem, an isolated variable experiment was performed. In detail, the NBCS constraint is removed from

TABLE IV. Study of different initial feature-matching methods.

Matching method	Precision (%)	Recall (%)	f-score (%)	Time (s)	ME (pixels)	RMSE (pixels)
SIFT	96·84	97·34	97·07	1·16	3·18	1·28
PSO-SIFT	94·79	99·05	96·79	1·25	3·15	1·31
SAR-SIFT	96·38	97·18	96·72	1·21	3·21	1·38

TABLE V. Isolated variable experiment. Note that NBCS⁻ preserves many outliers with large residuals; thus, its ME and RMSE are also very large – values greater than 50 pixels are indicated with an asterisk (*).

Method	Precision (%)	Recall (%)	f-score (%)	Time (s)	ME (pixels)	RMSE (pixels)
NBCS⁻	86·15	70·66	73·68	2·34	*	*
NBCS⁺	96·84	97·34	97·07	1·16	3·18	1·28

(a) Initial SIFT matches.

(b) LLT result.

(c) VFC result.

(d) RANSAC result.

(e) MLESAC result.

(f) AC-RANSAC result.
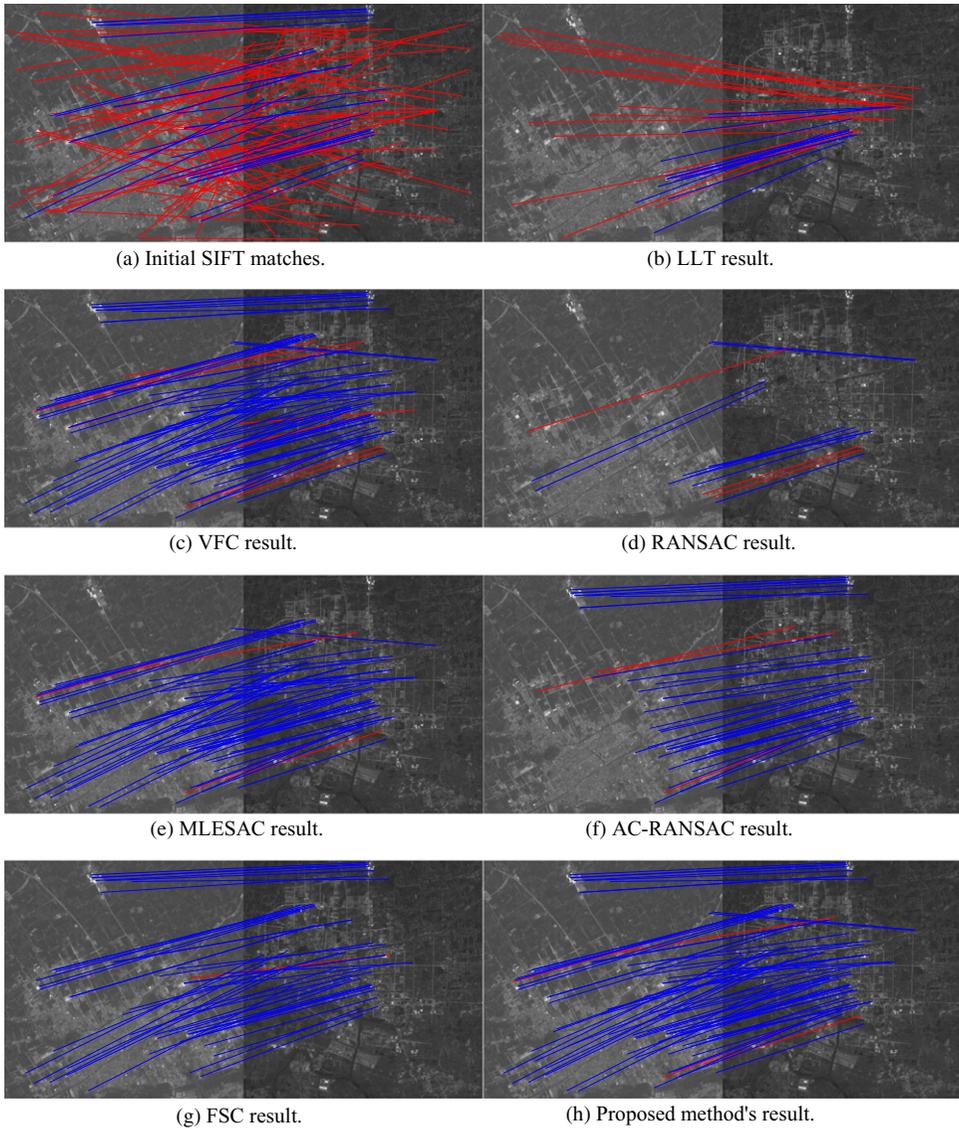
(g) FSC result.

(h) Proposed method's result.

Fig. 4. Qualitative comparisons for image pair 8 (SPOT-5 satellite images). The blue and red lines represent correct and false matches, respectively. For better visualisation, no more than 100 randomly selected matches are presented.

Algorithm 1 (denoted by NBCS⁻) and NBCS⁻ is performed on the dataset. Next, the NBCS constraint is added (denoted by NBCS⁺) and the experiment is performed again. The consequent precision, recall, f-score, running time, ME and RMSE metrics are reported in Table V. As can be seen, NBCS⁺ significantly outperforms NBCS⁻, which gains 23·4% in terms of f-score compared with NBCS⁻. Since NBCS⁻ preserves many false matches with

large residuals in their results, the ME and RMSE of NBCS⁻ are very large. In addition, with NBCS constraints, the number of transformation estimation times in Algorithm 1 can be largely reduced. Thus, the running time of NBCS⁺ is less than NBCS⁻.

Results and Discussion

Qualitative Comparisons. Fig. 4 illustrates the qualitative results on image pair 8 (SPOT-5 images near Shanghai, China). This pair has an initial inlier rate of 8·94%, which suffers from severe temporal changes (the component images were taken four years apart) and rotational changes. As shown, most detected matches using LLT are outliers; as a result, its precision is very low. RANSAC only extracts a few true inliers, most of which are filtered as outliers; thus, the recall accuracy is very low. Although VFC achieves a very impressive result on this image pair, it preserves many more outliers than the proposed method. MLESAC, AC-RANSAC and FSC achieve similar results to the proposed method on this image pair.

Fig. 5 reports the qualitative comparisons on image pair 13 (false-colour infrared aerial photographs over Illinois, USA) with an initial inlier rate of 1·6%, which suffers from an extremely small overlapping region. As can be seen, LLT, VFC, MLESAC, AC-RANSAC and FSC have totally failed. RANSAC performs much better; however, it still preserves some gross errors. The proposed method achieves the best performance.

Quantitative Evaluation. The quantitative comparisons are reported in Table VI. The average inlier rate of these 20 image pairs is 12·62%. In other words, the initial matches provided by SIFT contain a massive average outlier rate of 87·38%. Matching with this particular data is very difficult due to the extremely low inlier rate (14 of the 20 image pairs have an inlier rate lower than 10%).

As can be seen, LLT and VFC perform poorly on this dataset. The f-score for these two techniques is lower than 70%. This may be expected if their methodologies are assessed. VFC is a non-parametric method which does not estimate the transformation. As a result, noisy correspondences with relatively low locational accuracy may not be distinguished from gross errors effectively. LLT is an iterative method based on EM, which may not converge to a correct solution with challenging image pairs. The recall accuracy of AC-RANSAC is very low. RANSAC and MLESAC perform much better than LLT and VFC; however, they still preserve many outliers with their f-score being about 80%. They estimate the transformation parameters by closed-form methods, which are sensitive to noise and the estimated transformation may be skewed. FSC uses a small distance ratio to sample the initial feature correspondences and estimates affine transformation on the sampling set, which improves the performance compared with the traditional RANSAC method. The proposed method achieves the best accuracy on both the precision and recall metrics, meaning the method achieves the highest f-score. The proposed method uses the NBC principle to select subsets for transformation estimation, which largely improves robustness and reduces the transformation estimation trails compared with traditional hypothesise-and-verify methods such as RANSAC. As shown, the proposed method gains 28·7%, 27·9%, 17·0%, 15·6%, 26·5% and 8·8%, in terms of f-score, compared with LLT, VFC, RANSAC, MLESAC, AC-RANSAC and FSC, respectively. The average ME and RMSE of the proposed method are 3·18 and 1·28 pixels, respectively. There are no gross errors in the results because the outliers in the proposed method are correspondences with relatively low precision. In addition, the running time of the proposed method is only slightly higher than LLT (the method with the shortest time). It is the third fastest (note that the fastest method,

Li et al. Robust feature matching for geospatial images via an affine-invariant coordinate system

(a) Initial SIFT matches. (b) LLT result. (c) VFC result. (d) RANSAC result.

(e) MLESAC result. (f) AC-RANSAC result. (g) FSC result. (h) Proposed method.
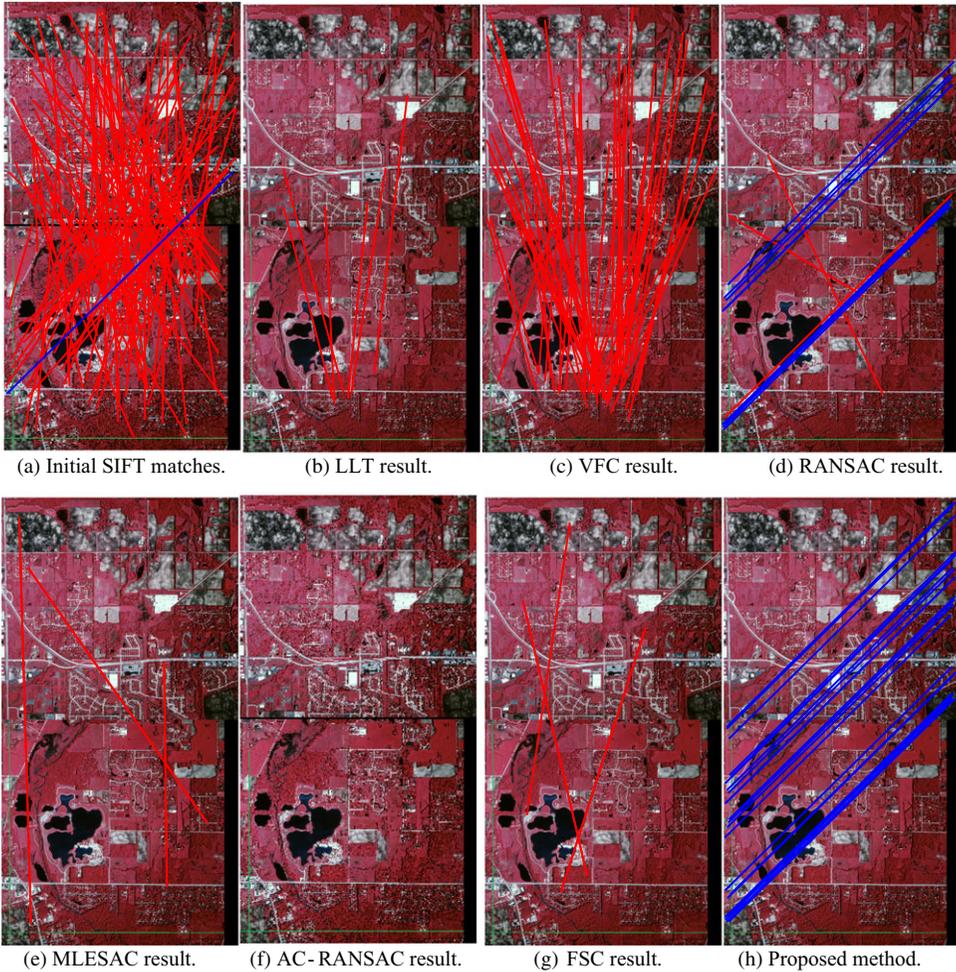
Fig. 5. Qualitative comparisons of image pair 13 (aerial photographs). The blue and red lines represent correct and false matches, respectively. For better visualisation, no more than 100 randomly selected matches are presented.

Table VI. Quantitative evaluation of the different feature-matching methods over the 20 datasets. Best results are shown in bold.

Metric	Proposed	LLT	VFC	RANSAC	MLESAC	AC-RANSAC	FSC
Precision (%)	**96·84**	63·85	61·75	88·18	91·19	83·27	95·87
Recall (%)	**97·34**	74·24	79·77	75·87	76·69	68·74	86·81
f-score (%)	**97·07**	68·32	69·16	80·03	81·45	70·57	88·96
Time (s)	1·16	**1·01**	15·95	36·21	5·21	**0·24 (C++)**	2·07

AC-RANSAC, is implemented using C++) of these seven methods. In the proposed method, a non-uniform sampling strategy and the NBC principle are adopted to reduce the computational complexity significantly.

Limitations of the Proposed Method. There are two main drawbacks to the proposed method:

(1) The running time may be affected by the initial inlier rates. The lower the initial inlier rate, the higher the time cost.
(2) The invariance of NBC will be not satisfied by non-rigid or close-range image pairs. Thus, the proposed method is not suitable for all types of image matching.

Conclusion

In this paper, an affine-invariant coordinate system called the normalised barycentric coordinate system (NBCS) is defined, which is superior to the Cartesian coordinate system for robust image matching and registration. The defined coordinate system has been adapted into a hypothesise-and-verify framework. The qualitative and quantitative comparisons of 20 typical geospatial image pairs demonstrate the power of the proposed method, which significantly outperforms six other state-of-the-art methods (LLT, VFC, RANSAC, MLESAC, AC-RANSAC and FSC).

Acknowledgements

References

CHO, M., SUN, J., DUCHENNE, O. and PONCE, J., 2014. Finding matches in a haystack: a max-pooling strategy for graph matching in the presence of outliers. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2014)*, Columbus, Ohio, USA. 2091–2098.

CHUM, O. and MATAS, J., 2005. Matching with PROSAC – progressive sample consensus. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2005)*, San Diego, California, USA. 1: 220–226.

CHUM, O., MATAS, J. and KITTLER, J., 2003. Locally optimized RANSAC. *Joint Pattern Recognition Symposium (DAGM 2003)*, Magdeburg, Germany. 8: 236–243.

DELLINGER, F., DELON, J., GOUSSEAU, Y., MICHEL, J. and TUPIN, F., 2015. SAR-SIFT: a SIFT-like algorithm for SAR images. *IEEE Transactions on Geoscience and Remote Sensing*, 53(1): 453–466.

DEMPSTER, A. P., LAIRD, N. M. and RUBIN, D. B., 1977. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, 39(1): 1–38.

FISCHLER, M. A. and BOLLES, R. C., 1981. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6): 381–395.

HARTLEY, R. and ZISSERMAN, A., 2003. *Multiple View Geometry in Computer Vision*. Second edition. Cambridge University Press, Cambridge, UK. 655 pages.

HAST, A., NYSJÖ, J. and MARCHETTI, A., 2013. Optimal RANSAC – towards a repeatable algorithm for finding the optimal set. *Journal of WSCG*, 21(1): 21–30.

HU, H., ZHU, Q., DU, Z., ZHANG, Y. and DING, Y., 2015. Reliable spatial relationship constrained feature point matching of oblique aerial images. *Photogrammetric Engineering & Remote Sensing*, 81(1): 49–58.

LEBEDA, K., MATAS, J. and CHUM, O., 2012. Fixing the locally optimized RANSAC – full experimental evaluation. *British Machine Vision Conference (BMVC 2012)*, Guildford, UK. 11 pages.

LI, B. and YE, H., 2012. RSCJ: robust sample consensus judging algorithm for remote sensing image registration. *IEEE Geoscience and Remote Sensing Letters*, 9(4): 574–578.

Li, J., Hu, Q. and Ai, M., 2016. Robust feature matching for remote sensing image registration based on L_q-estimator. *IEEE Geoscience and Remote Sensing Letters*, 13(12): 1989–1993.

Li, X. and Hu, Z., 2010. Rejecting mismatches by correspondence function. *International Journal of Computer Vision*, 89(1): 1–17.

Liu, Z., An, J. and Jing, Y., 2012. A simple and robust feature point matching algorithm based on restricted spatial order constraints for aerial image registration. *IEEE Transactions on Geoscience and Remote Sensing*, 50(2): 514–527.

Lowe, D. G., 2004. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2): 91–110.

Ma, J., Zhao, J., Tian, J., Yuille, A. L. and Tu, Z., 2014. Robust point matching via vector field consensus. *IEEE Transactions on Image Processing*, 23(4): 1706–1721.

Ma, J., Zhou, H., Zhao, J., Gao, Y., Jiang, J. and Tian, J., 2015. Robust feature matching for remote sensing image registration via locally linear transforming. *IEEE Transactions on Geoscience and Remote Sensing*, 53(12): 6469–6481.

Ma, W., Wen, Z., Wu, Y., Jiao, L., Gong, M., Zheng, Y. and Liu, L., 2017. Remote sensing image registration with modified SIFT and enhanced feature matching. *IEEE Geoscience and Remote Sensing Letters*, 14(1): 3–7.

Moisan, L., Moulon, P. and Monasse, P., 2012. Automatic homographic registration of a pair of images, with *a contrario* elimination of outliers. *Image Processing On Line*, 2: 56–73.

Raguram, R., Chum, O., Pollefeys, M., Matas, J. and Frahm, J.-M., 2013. USAC: a universal framework for random sample consensus. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(8): 2022–2038.

Sedaghat, A., Mokhtarzade, M. and Ebadi, H., 2011. Uniform robust scale-invariant feature matching for optical remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 49(11): 4516–4527.

Torr, P. H. S. and Zisserman, A., 2000. MLESAC: a new robust estimator with application to estimating image geometry. *Computer Vision and Image Understanding*, 78(1): 138–156.

Torresani, L., Kolmogorov, V. and Rother, C., 2008. Feature correspondence via graph matching: models and global optimization. *10th European Conference on Computer Vision (ECCV 2008)*, Marseille, France. 596–609.

Vedaldi, A. and Fulkerson, B., 2010. VLFeat: an open and portable library of computer vision algorithms. *Proceedings of the 18th ACM International Conference on Multimedia (ACM MM, 2010)*, Florence, Italy. 1469–1472.

Wu, Y., Ma, W., Gong, M., Su, L. and Jiao, L., 2015. A novel point-matching algorithm based on fast sample consensus for image registration. *IEEE Geoscience and Remote Sensing Letters*, 12(1): 43–47.

Yang, Z. and Cohen, F. S., 1999. Image registration and object recognition using affine invariants and convex hulls. *IEEE Transactions on Image Processing*, 8(7): 934–946.

Zhang, K., Li, X. and Zhang, J., 2014. A robust point-matching algorithm for remote sensing image registration. *IEEE Geoscience and Remote Sensing Letters*, 11(2): 469–473.

Résumé

L'appariement de caractéristiques est une étape cruciale pour de nombreuses applications de la photogrammétrie et de la télédétection telles que l'enregistrement d'images et la reconstruction 3D. Dans cet article, nous proposons une méthode d'appariement de caractéristiques nouvelle et robuste basée sur le système de coordonnées barycentriques normalisées (NBCS), qui est supérieur au système de coordonnées cartésiennes pour cette tâche. Nous effectuons une transformation d'entité invariante par changement d'échelle (SIFT) pour fournir des correspondances initiales contenant à la fois des correspondances correctes (inliers) et des correspondances fausses (outliers), en insistant sur l'estimation de modèle à partir d'observations contaminées (corresponsances fausses). Nous définissons un système de coordonnées invariant par transformation affine appelé NBCS basé sur des rapports de surfaces. Les deux points caractéristiques d'une correspondance correcte ont les mêmes coordonnées sous NBCS, contrairement aux correspondances fausses. Nous adaptons ce principe selon une approche par hypothèse et vérification. La méthode proposée est robuste, efficace et efficiente. Des expériences approfondies sur des paires d'images réelles de télédétection montrent qu'elle surpasse significativement les approches de l'état de l'art. Tous les codes sources et jeux de données utilisés dans le document sont rendus publics.

Zusammenfassung

Die Merkmalsgestützte Bildzuordnung ist ein wichtiger Schritt für viele photogrammetrische Anwendungen wie Bildregistrierung und 3D-Rekonstruktion. In diesem Beitrag schlagen wir eine neuartige und robuste Merkmalsgestützte Bildzuordnungsmethode auf der Grundlage eines normalisierten Baryzentrischen Koordinatensystems (NBCS) vor, das einem kartesischen Koordinatensystem für diese Aufgabe überlegen ist. Zunächst wird eine skaleninvariante Merkmalstransformation (SIFT) durchgeführt, um erste Zuordnungen zu liefern, die sowohl korrekte Zuordnungen als auch falsche Zuordnungen (Ausreißer) enthalten. Dann wird auf die Modellschätzung von kontaminierten Beobachtungen (Übereinstimmungen mit Ausreißern) fokussiert. Es wird ein affin-invariantes Koordinatensystem namens NBCS auf der Grundlage von Verhältnissen von Bereichen eingeführt. Zwei homologe Merkmalspunkte einer korrekten Übereinstimmung haben die gleichen Koordinaten im NBCS Koordinatensystem, während falsche Korrespondenzen diese Eigenschaft nicht besitzen. Dieses Prinzip wird in eine Hypothesen-und-Verifizierungs-Struktur eingebaut. Die vorgeschlagene Methode ist robust, effizient und effektiv. Umfangreiche Experimente an realen Bildpaaren zeigen, dass die vorgeschlagene Methode sechs andere, moderne Ansätze deutlich übertrifft. Der gesamte Quellcode und der Datensatz, der in diesem Beitrag verwendet wird, sind öffentlich zugänglich.

Resumen

La correspondencia de imágenes es una etapa crucial en muchas aplicaciones de fotogrametría y teledetección, como el registro de imágenes y la reconstrucción en 3D. En este artículo proponemos un nuevo y robusto método de correspondencia basada en características que se basa en un sistema de coordenadas baricéntrico normalizado (NBCS), que es superior al sistema de coordenadas cartesianas para esta tarea. Se aplica el descriptor SIFT para proporcionar correspondencias iniciales que contienen tanto correspondencias correctas (inliers) como correspondencias falsas (outliers), y con un enfoque en la estimación del modelo a partir de observaciones contaminadas (correspondencias con outliers). Se define, en base a la razón de áreas, un sistema de coordenadas afín-invariante (NBCS). Los dos puntos de una correspondencia correcta tienen la misma coordenada en el sistema NBCS mientras que las correspondencias falsas no cumplen esta propiedad. Adaptamos este principio a un marco de hipótesis y verificación. El método propuesto es robusto, eficiente y efectivo. Los experimentos extensivos sobre pares de imágenes geoespaciales reales muestran que el método propuesto supera significativamente otros métodos del estado del arte actual. Todo el código fuente y el conjunto de datos utilizados en el documento se han hecho públicos.

摘要

特征匹配是许多遥感应用的关键步骤。本文提出一种新颖鲁棒的基于标准化重心坐标系(NBCS)的特征匹配方法,该坐标系相比于笛卡尔坐标系更适合于特征匹配任务。我们采用SIFT算法提供包含正确匹配(内点)和错误匹配(外点)的初始匹配对,并且关注于从被污染的观测值中(包含粗差)估计模型。我们基于面积比定义了一种仿射不变坐标系,称为NBCS。正确匹配的两个特征点在NBCS下具有相同的坐标,而错误匹配点不遵循该特性。我们将这一原理应用于假设-检验框架下。所提出的方法是鲁棒的,快速的和有效的,即它在我们的数据集。对真实遥感图像对的广泛实验显示其显著优于其他先进方法。本文中使用的所有源代码和数据集都是公开的。