Multiview Matching Algorithm for Processing Mobile Sequence Images

Xiaoliang Zou, Ph.D.[1]; Guihua Zhao[2]; Jonathan Li, Ph.D., P.Eng.[3]; Yuanxi Yang, Ph.D.[4]; and Yong Fang, Ph.D.[5]

Abstract: The paper presents a multiview matching algorithm for processing sequence images acquired by a mobile mapping system (MMS). The workflow of the multiview matching algorithm is designed, and the algorithm is based on motion analysis of sequence images in computer vision. To achieve a high multiview matching accuracy, camera lens distortion in sequence images is first corrected, and images can then be resampled. Image points on sequence images are extracted using the Harris operator. The homologous image points are then matched based on correlation coefficients and used to make a robust estimation for a fundamental matrix F between the two adjacent images using the random sample consensus (RANSAC) algorithm. The fundamental matrix F is calculated under the condition of epipolar line constraints. Finally, the trifocal tensor T of the three-view images is calculated to achieve highly accurate triplet image points. These triplet image points are then provided as the initial value for bundle adjustment. The algorithm was tested using a set of sequence images. The results demonstrate that the designed workflow is available and the algorithm is promising in terms of both accuracy and feasibility. **DOI: 10.1061/(ASCE)SU.1943-5428.0000235.**

Author keywords: Sequence images; Multiview matching; Harris operator; Correlation coefficient; Random sample consensus (RANSAC); Trifocal tensor; Computer vision (CV); Mobile mapping system (MMS).

Introduction

Photogrammetry and geometric computer vision are closely related disciplines. Many studies have shared interest in these two disciplines for many similar goals, such as point feature detection (Forstner and Gulch 1987), relative orientation (Philip 1996; Nister 2004), perspective n-point (PnP) problems (Masry 1981; Lepetit et al. 2009), and bundle adjustment (Triggs et al. 2000). The mathematical fundamentals of photogrammetry and computer vision can both be derived from the central projection of the common mathematical model. Photogrammetry uses the collinearity equations of Cartesian coordinate representation of the central projection in Euclidean geometry, and computer vision applies the projection equations of homogeneous coordinate representation of the central projection in projective geometry. Homogenous coordinates have the advantage that the points, lines, and planes at infinity can be represented using finite coordinates. In photogrammetry, the nonlinearity of the collinearity equations requires linearity and iterative optimization and good initial values of exterior orientation (EO) parameters from a global navigation satellite system and inertial measuring unit (GNSS/IMU) system. In addition, all the parameters in collinearity equations have physical meanings. In computer vision, the linearity of the projection equation permits linear matrix operations using linear algebra, the linearity of the camera matrix or fundamental matrix does not require the initial values, and those parameters of the matrix are not physically interpretable.

Multiview image matching has been addressed by several researchers in photogrammetric and computer vision. Maas (1996) presented a multi-image matching algorithm using discrete points extracted by an interest operator and epipolar line intersection. Brown et al. (2005) presented a method of multi-image matching in which image features are first located as interest points using a Harris corner detector, followed by matching using a fast nearest neighbor algorithm that indexes features based on their low-frequency Haar wavelet coefficients, followed by refining feature matches using the random sample consensus (RANSAC) method. Gruen (1985) presented a method of multiphoto correlation based on the geometrically constrained adaptive least-squares matching algorithm. Elaksher (2008) presented a method of using forward neural networks to solve multi-image correspondence and using the photogrammetric collinearity condition to validate the outputs of the neural network and to compute the three-dimensional (3D) coordinates of the matched points.

With the rapid development of mobile mapping system (MMS) technology in recent years, real-time sequence images of motion can be collected by digital cameras mounted on a platform of the vehicle-based mobile mapping system. The sequence images acquired from a mobile mapping system can be matched using multiple views based on the methods of computer vision. This paper presents a multiview matching algorithm based on motion analysis of sequence images that uses the well-known algorithms of

[1] Senior Engineer, State Key Laboratory of Geo-Information Engineering, Xi'an 710054, P. R. China (corresponding author). ORCID: https://orcid.org/0000-0002-4561-6701. E-mail: x26zou@uwaterloo.ca

[2] Engineer, Xi'an Institute of Surveying and Mapping, Xi'an 710054, P. R. China. E-mail: zghzhz2010@163.com

[3] Professor, Dept. of Geography and Environmental Management, Faculty of Environment, Univ. of Waterloo, Waterloo, ON, Canada N2L 3G1. E-mail: junli@uwaterloo.ca

[4] Professor, State Key Laboratory of Geo-Information Engineering, Xi'an 710054, P. R. China. E-mail: yuanxi_yang@163.com

[5] Researcher, Xi'an Institute of Surveying and Mapping, Xi'an 710054, P. R. China. E-mail: yong.fang@vip.sina.com

Note. This manuscript was submitted on March 9, 2016; approved on March 9, 2017; published online on August 17, 2017. Discussion period open until January 17, 2018; separate discussions must be submitted for individual papers. This paper is part of the ***Journal of Surveying Engineering***, © ASCE, ISSN 0733-9453.

computer vision to process sequence images rapidly and get highly accurate image point coordinates.

Design Ideas and Workflow

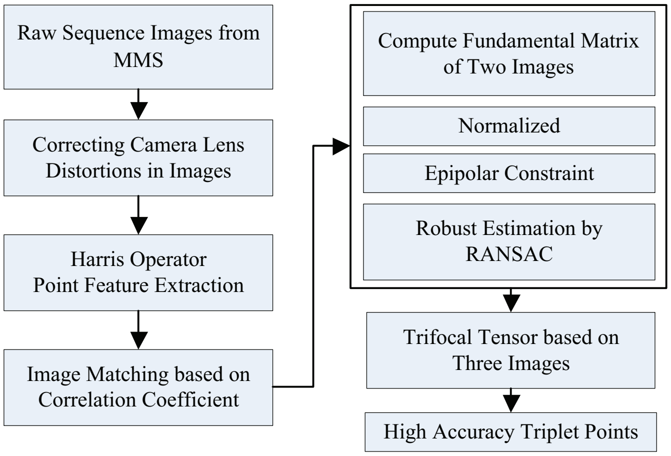
Fig. 1 presents the workflow diagram of the proposed method. The method is based on motion analysis of sequence images using the well-known algorithms of computer vision. To achieve higher matching accuracy, camera lens distortions in sequence images are first corrected, and images can then be resampled. Image points on each nondistortion sequence image are extracted using the Harris operator. Then, the homologous image points are matched based on correlation coefficients. The matched homologous points are then used to fit the fundamental matrix F between the two adjacent images using the RANSAC algorithm for robust estimation. A fundamental matrix F of two views is calculated under the condition of epipolar line constraints. Then, the trifocal tensor T of the three-view images is calculated to achieve highly accurate triplet image points. To test the algorithm, a set of sequence images was captured using a Sony (Tokyo, Japan) DFW-SX910 camera in a MMS, and the experimental results were analyzed.

Sequence-Image Matching Algorithm

The Applanix (Richmond Hill, Ontario, Canada) Landmark MMS used in this study consisted of charge-coupled device (CCD) digital cameras, laser scanner sensors, a GNSS/IMU positioning and orientation system, and an odometer. The CCD digital cameras and laser scanners were mounted on the top of the land vehicle. The CCD cameras pointed in different directions to acquire different digital images. The front-view sequence images were used as test samples in this study. The proposed algorithm of sequence-image matching consists of camera lens distortions correction, use of Harris operator to extract interest points, two-view image matching based on correlation coefficient, fitting of fundamental matrix F of two views by RANSAC, and estimate of trifocal tensor T of three views. The algorithm is detailed in the following sections.

Camera Lens Distortions Correction

Camera lens distortions cause the imaged positions to be displaced from their ideal location. Lens distortions are the origin for systematic errors in sequence-image coordinates. The metric digital camera sensors mounted on the MMS were rigorously calibrated to effectively correct systematic errors in image points. Camera calibration parameters are provided to the user in camera calibration report, including focal length f, principal point shift (x_0, y_0) parameters, radial distortion parameters (k_1, k_2, k_3) and tangential distortion parameters(p_1, p_2).



Raw Sequence Images from MMS
Correcting Camera Lens Distortions in Images
Harris Operator Point Feature Extraction
Image Matching based on Correlation Coefficient
Compute Fundamental Matrix of Two Images
Normalized
Epipolar Constraint
Robust Estimation by RANSAC
Trifocal Tensor based on Three Images
High Accuracy Triplet Points

Fig. 1. Workflow diagram of multiviews matching algorithm

With known camera lens distortion parameters, the sequence-image coordinates may be refined more effectively with the Brown distortion model (Brown 1971, 1976). The corrected coordinates $(\bar{x}, \bar{y})$ are computed with Eqs. (1)–(4).

$$\begin{aligned}\bar{x} &= x - x_0 + \Delta x\\ \bar{y} &= y - y_0 + \Delta y\end{aligned} \tag{1}$$

where

$$\begin{aligned}\Delta x &= \Delta x_1 + \Delta x_2\\ \Delta y &= \Delta y_1 + \Delta y_2\end{aligned} \tag{2}$$

Radial distortion:

$$\begin{aligned}\Delta x_1 &= (x - x_0)\left(k_1 r^2 + k_2 r^4 + k_3 r^6\right)\\ \Delta y_1 &= (y - y_0)\left(k_1 r^2 + k_2 r^4 + k_3 r^6\right)\end{aligned} \tag{3}$$

Tangential distortion:

$$\begin{aligned}\Delta x_2 &= p_1\left[r^2 + 2(x - x_0)^2\right] + 2p_2(x - x_0)(y - y_0)\\ \Delta y_2 &= p_2\left[r^2 + 2(y - y_0)^2\right] + 2p_1(x - x_0)(y - y_0)\end{aligned} \tag{4}$$

where $(\bar{x}, \bar{y})$ = corrected image coordinate; (x, y) = raw image coordinate; $(\Delta x, \Delta y)$ = correction terms for formulating the camera's systematic error; and $r^2 = (x - x_0)^2 + (y - y_0)^2$.

The image coordinates of the sequence images from the MMS are corrected with the Brown distortion model. Raw sequence images are resampled, and new nondistortion images are formed for multiview matching.

Harris Operator

The Harris operator (Harris and Stephens 1988) is a combined corner and edge detector with geometric stability. It defines interest points that have locally maximal self-matching precision under translational least-squares template matching. The authors applied the Harris operator to extract interest points in the sequence images from the MMS. The Harris operator results from shifting a local window in the image by a small amount in various directions.

Denoting the image intensities with I in the x- and y-directions, the change E (Harris and Stephens 1988) produced by a shift (x, y) is given as

$$\begin{aligned}E_{x,y} &= \sum_{u,v} W_{u,v}\left(I_{x+u,y+v} - I_{u,v}\right)^2\\ &= \sum_{u,v} W_{u,v}\left[xX + yY + O\left(x^2, y^2\right)\right]^2\end{aligned} \tag{5}$$

where $W_{u,v}$ is weighted by a Gaussian and denotes

$$W_{u,v} = exp - \left(u^2 + v^2\right)/2\sigma^2 \tag{6}$$

The first gradients are approximated as

$$\begin{cases} X = I \otimes (-1, 0, 1) \approx \partial \mathrm{I}/\partial \mathrm{x}\\ Y = I \otimes (-1, 0, 1)^T \approx \partial \mathrm{I}/\partial y\end{cases} \tag{7}$$

For the small shift (x, y), the change E can be written as

$$E(x,y) = (x,y)M(x,y)^T = Ax^2 + 2Cxy + By^2 \tag{8}$$

Where the 2 × 2 symmetric matrix M is rewritten in Eq. (9) and the parameters denoted in Eq. (10), as follows:

$$M = \begin{bmatrix} A & C \\ C & B \end{bmatrix} \tag{9}$$

$$\begin{aligned} A &= X^2 \otimes W \\ B &= Y^2 \otimes W \\ C &= XY \otimes W \end{aligned} \tag{10}$$

where E is closely related to the local autocorrelation function; and M describes its shape at the origin. Let λ_1, λ_2 be the eigenvalues of M. λ_1 and λ_2 will be proportional to the principal curvatures of the local autocorrelation function and form a rotationally invariant description of M. The Harris corner region (Harris and Stephens 1988) is defined as shown in Eq. (11), with Det(M) and Tr(M) as shown in Eqs. (12) and (13), respectively, as follows:

$$R = \mathrm{Det}(M) - K \cdot \mathrm{Tr}(M)^2 \tag{11}$$

$$\mathrm{Det}(M) = \lambda_1\lambda_2 = AB - C^2 \tag{12}$$

$$\mathrm{Tr}(M) = \lambda_1 + \lambda_2 = A + B \tag{13}$$

R is positive in the corner region, negative in edge regions, and small in the flat region. The parameter K is usually set to 0.04–0.06. In Eqs. (12) and (13), an ideal edge is λ_1 large, λ_2 zero; a corner will be indicated by both λ_1 and λ_2 large; and a flat image region will be indicated by both λ_1 and λ_2 small. If $\lambda_1 \gg \lambda_2$, an edge is a horizontal edge; if $\lambda_1 \ll \lambda_2$, an edge is a vertical edge. Interest feature points are extracted in every sequence image using the Harris operator.

Sequence-Image Matching by Correlation Coefficient

Next, an image matching algorithm based on the correlation coefficient is applied to acquire homologous feature points. Image matching is implemented between previously detected interest feature points in two sequence images by finding points that are maximally correlated with each other within windows surrounding each point. Only interest points that correlate most strongly with each other in both the left–right and right–left directions are recorded as homologous matching points. The validity of homologous matching points can be checked in both directions.

The implementation step of image matching is as follows. First, the initial point sets detected by the Harris operator are divided into adjacent sequence images to form multipoint sets. Then, by correlation coefficient computation between a characteristic point in the match window and the candidate matching points within the search window point by point, the points with maximal correlation coefficients in both directions are regarded as homologous pairs of points. The correlation coefficient is used to estimate the similarity of gray vector linear correction and is an important similar-measurement method. The correlation coefficient between point $p(x,y)$ in the A frame image and point $q\ (x',y')$ in $A+1$ frame image is defined as

$$\rho = \frac{1}{\sigma_1\sigma_2}\sum_{j=-n}^{n}\sum_{i=-n}^{n}[I_A(x+i,y+j) - \mu_1][I_{A+1}(x'+i,y'+j) - \mu_2] \tag{14}$$

where $I_A(x,y)$ and $I_{A+1}(x',y')$ = gray intensities in the matching window and research window of the A frame image or the $A+1$ frame image, respectively; μ_1, μ_2, σ_1 and σ_2 = mean and standard deviation of the match windows and research windows, respectively; n = an arrange within small region. Of course, the right window size must be chosen to improve the speed and time of matching. Window size w should be odd, such as 11 × 11, 13 × 13. The radius of matching window is $(w-1)/2$. Finally, the correlation coefficient between sequence images is calculated according to Eq. (14), and the homologous pairs of points with maximal correlation in both directions are recorded.

Fit Fundamental Matrix of Two Views by RANSAC

For a given point in one sequence image, a corresponding point in the other sequence image may not exist. As a result of missing parts in images, mismatching, or lack of a sufficiently textured image, the matching algorithm may fail to find the homologous point and produce outliers. Thus, these outliers must be removed using the fundamental matrix F by estimating the epipolar geometry and RANSAC strategy to fit the fundamental matrix according to the matched points. The proposed method applies the epipolar geometry constraint and the well-known normalized 8-point algorithm to calculate the fundamental matrix F and fit the fundamental matrix of two views with the RANSAC algorithm.

Epipolar Geometry Constraint

The epipolar geometry is the intrinsic projective geometry between two views. Epipolar geometry is independent of scene structure and only depends on the camera's internal parameters and relative pose (Hartley and Zisserman 2000). In computer vision, the most common way to represent this intrinsic geometry is the fundamental matrix F, which is a 3 × 3 matrix of rank 2. A 3D-space point $M(X,Y,Z)$ is imaged as $m(x,y,1)^T$ in the first frame view and $m\prime(x\prime,y\prime,1)^T$ in the second frame view, and then the image points satisfy the relation $m'^T Fm = 0$; that is, the camera center, 3D-space point M, and its image points m and m' lie in a common plane.

The fundamental matrix F can be derived from the mapping between a point and its epipolar line. For a given point m in the first frame image, the projective representation of the epipolar line in the second frame image l' is given by $l' = Fm$; the point m' corresponding to m belongs to the line l', and therefore, $m'^T l' = m'^T Fm = 0$. Thus, the fundamental matrix F satisfies the condition that for any pair of corresponding points $m \leftrightarrow m'$ in those two frame images, $m'^T Fm = 0$.

Normalized 8-Point Algorithm

The simplest method of computing the fundamental matrix is an 8-point algorithm, as presented by Longuet-Higgins (1981). Given sufficiently many point matches $m \leftrightarrow m'$ (at least 8 points), the equation $m'^T Fm = 0$ can be used to compute the unknown matrix F. Each point and its corresponding matched point give rise to one linear equation in the unknown entries of F. According to the known coordinates m and m', the equation corresponding to a pair of point m and point m' can be expressed as a vector inner product by

$$(x'x,\ x\prime y,\ x',\ y\prime x,\ y\prime y,\ \ y\prime,\ x,\ y,1)f\ =\ 0 \tag{15}$$

where $f = [\,F_1 \quad F_2 \quad F_3 \quad F_4 \quad F_5 \quad F_6 \quad F_7 \quad F_8 \quad F_9\,]$ is a 9-point vector made up of the entries of fundamental matrix F in row-major order. For a set of $n\ \geq\ 8$ matched points, a set of linear equations of the following form is obtained:

$$Bf = \begin{bmatrix} x'_1x_1 & x'_1y_1 & x'_1 & y'_1x_1 & y'_1y_1 & y'_1 & x_1 & y_1 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x'_nx_n & x'_ny_n & x'_n & y'_nx_n & y'_ny_n & y'_n & x_n & y_n & 1 \end{bmatrix} f = 0 \tag{16}$$

For a solution to exist, matrix B must have ranked at most 8. But the rank of B may be greater than 8 because of noise in the point coordinates and a lack of exact data. In this case, the least-squares solution must be found. The least-squares solution for f is the singular vector corresponding to the smallest singular value of B, that is, the last column of V in the SVD $B = \mathrm{UDV}^T$. The solution vector f found in this way minimizes $||Bf||$ subject to the condition $||f|| = 1$.

Table 1. DFW-SX910 9833406 Camera Parameters

Parameter	Value
Focal (mm)	−6.092963
x_o (mm)	−0.001046
y_o (mm)	−0.005866
k_1	0.002551
k_2	0.00
k_3	0.00
p_1	−0.000066
p_2	0.00
CCD width (pixel)	1,280.00
CCD height (pixel)	960.00
X pixel size (μm)	4.65
Y pixel size (μm)	4.65

The key to success with the 8-point algorithm is proper careful normalization of input data before constructing the equation. A simple transformation of the points in the image before formulating the linear equations leads to an enormous improvement in the conditioning of the problem and hence in the stability of the result. Normalization is a translation and scaling of each set of points in the image so that the origin of the reference points is at centroid, the mean distance of the points from the origin is equal to $\sqrt{2}$, and the scale parameter is 1. The normalized 8-point algorithm is a method for enforcing the singularity constraint on the fundamental matrix F. The initial estimate F is replaced by the singular matrix $\widehat{F}$ that minimizes the difference $||\widehat{F} - F||$ subject to the condition $\det \widehat{F} = 0$. This is done using the SVD, and has the advantage of being simple and rapid.

Detailed implementation of the step is as follows. First, transform the image coordinates according to $\widehat{x}_i = Tx_i$ and $\widehat{x}'_i = T'x'_i$, where T and T' are normalizing transformations. Second, compose matrix B from the pairs of matched homologous points $(x, y) \leftrightarrow (x', y')$ as defined in $Bf = 0$. Third, obtain a solution F from the vector f corresponding to the smallest singular value of B. Replace F with $\widehat{F}$, using the SVD for correction. Finally, set $F = T'^T\widehat{F}T$ for denormalization. Matrix F is the fundamental matrix corresponding to the original data.

Fit Fundamental Matrix by RANSAC

The RANSAC algorithm is used for the robust fitting of models in the presence of many data outliers (Fishler and Boles 1981). The

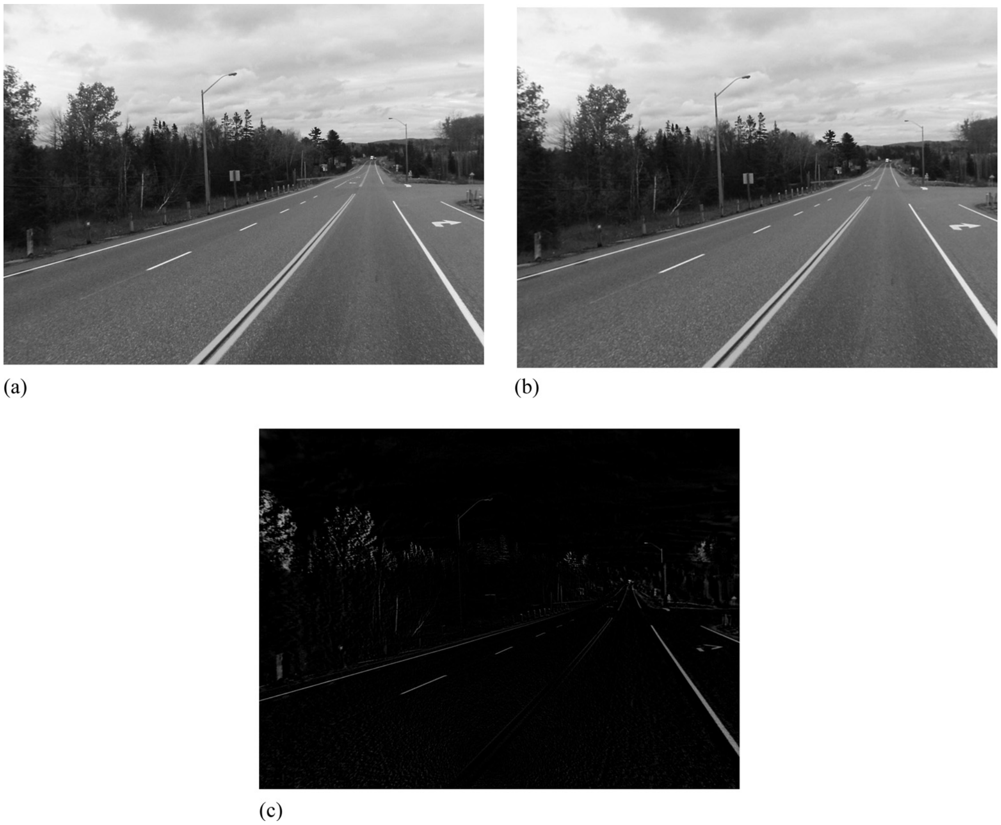


Fig. 2. (a) Raw front-view sequence image in Frame 149; (b) nondistortion sequence image in Frame 149; (c) difference image between (a) and (b) (images courtesy of Applanix Corporation)

algorithm is used to robustly fit a fundamental matrix to a set of putatively matched image points and obtain subset inliers (valid points). Given N matched pairs of the homologous point, the RANSAC algorithm is used to perform the iterative computations, as follows: (1) randomly select s sample correspondences as an initial data set. A minimum of s pairs of point numbers is required to form a fundamental matrix F and compute the fundamental matrix. Random sample s and the number of samples N creates the following relationship (Hartley and Zisserman 2000):

$$N = \log(1 - p)/\log\left[1 - (1 - \varepsilon)^{s}\right] \qquad (17)$$

Eq. (17) gives the number of samples N required to ensure, with a probability $p = 0.99$, that at least one sample has no outliers for a given size of sample s and proportion of outliers ε. (2) Compute the distance measure. Given a current estimate of F from the RANSAC sample, the distance d measures how closely a matched pair of points satisfies the epipolar geometry. (3) Compute the number of inliers consistent with F by the number of correspondences for which $d < t$. Value t is a distance threshold of F model for a 95% probability that the point is an inlier. Then, the solution for F with the most inliers is retained. (4) Choose the F with the largest number of inliers and re-estimate F for all correspondences classified as inliers. Thus, the best optimization of fundamental matrix F is fitted using the RANSAC robust estimation algorithm to remove the outliers.

Estimate Trifocal Tensor of Three Views

The trifocal tensor (Hartley and Zisserman 2003) encapsulates all the projective geometric relations between three views that are independent of scene structure. It only depends on the motion between views and the internal parameters of the camera. The trifocal tensor captures point–point–point correspondence between the three images. Corresponding points backprojected from each image all intersect in a single 3D point in space. A point in 3D-space is imaged as the corresponding triplet $x \leftrightarrow x' \leftrightarrow x''$ in three images. Because the three fundamental matrices F_{12}, F_{23}, and F_{34} relating the three views in sequence images are computed and homogeneous points between two views are matched, it is possible to determine the trifocal tensor given the three fundamental matrices and homogeneous points.

The methods of estimating trifocal tensor T are as follows: (1) Search three sets of corresponding image points from the matched homogeneous pairs among two views, and form the homogeneous point sets $x \leftrightarrow x' \leftrightarrow x''$ in three images. (2) Normalize each set of points so that the origin is at centroid, mean distance from origin is $\sqrt{2}$, and scale parameter is 1. (3) Randomly select more than 7

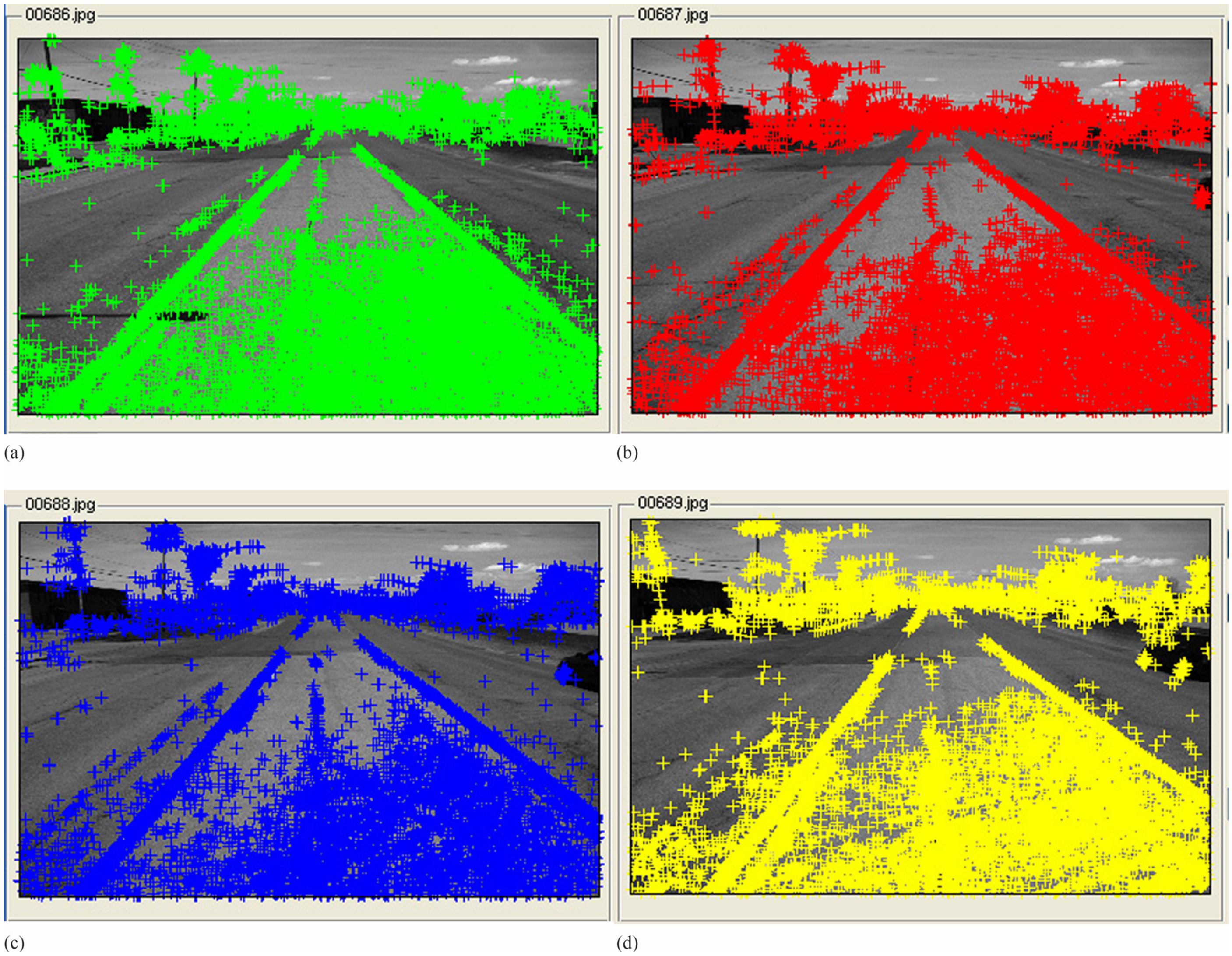


Fig. 3. Feature points in subplots: (a) Frame 686; (b) Frame 687; (c) Frame 688; (d) Frame 689

image point correspondences in the matched pairs of homologous points across the three images and form a fundamental matrix. (4) Compute the trifocal tensor T (Hartley and Zisserman 2003) using Eq. (18). On the condition of error constraints based on minimizing algebra and the homologous matched points of the three views, the trifocal tensor over the three images is estimated through the fundamental matrix using the RANSAC algorithm. Ensure the trifocal tensor is geometrically valid by retrieving its epipoles. (5) Perform denormalization and compute the trifocal tensor T to the original data.

$$x^i x'^j x''^k \epsilon_{jqs} \epsilon_{krt} T_i^{qr} = 0_{st} \tag{18}$$

where $x^i \leftrightarrow x'^j \leftrightarrow x''^k$ = a set of corresponding points across three frame views; ϵ = algebraic error vector and is the norm of the error vector that is minimized; T_i^{qr} = trifocal tensor; and 0_{st} = two-dimensional (2D) tensor with all zero entries.

The estimated trifocal tensor together with a set of corresponding triplet points $x \leftrightarrow x' \leftrightarrow x''$ across the three images is determined. The trifocal tensor is used to determine the exact image positions of three homologous points in three images. There are fewer mismatches over three views than there are over two views. There is only the weaker geometric constraint of an epipolar line against which to verify a possible match in the two views. Thus, the highly accurate and reliable results of the corresponding triplet points can be obtained. In future work, the authors will use image point coordinates from a set of triplet points in three views and EO parameters of the projection center from GNSS/IMU data postprocessing to solve a set of 3D points X by bundle adjustment over three views.

Results and Analysis

Data Set Description

The Sony DFW-SX910 9833406 camera is a front-view digital camera and was mounted in an Applanix Landmark MMS. The camera contains a digital CCD sensor with 1,280 × 960 pixels with a 4.65-μm pixel size, and the lens focal is 6.09 mm. The camera parameters are shown in Table 1. In this study, the authors chose the continuous adjacent four-frame sequence images from Frames 686, 687, 688, and 689 collected by the Sony DFW-SX910 9833406 camera as test samples.

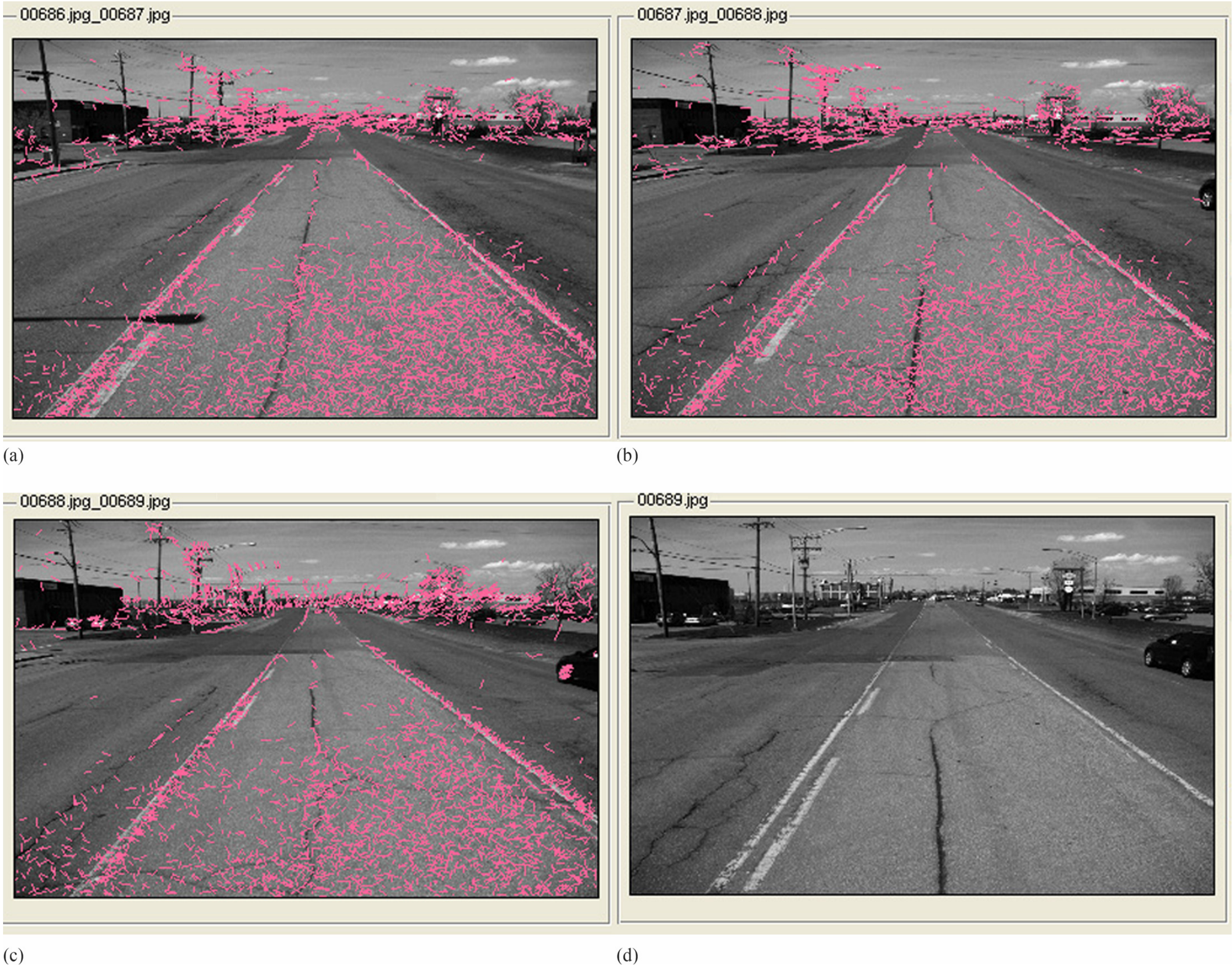


Fig. 4. Subplots (a) through (c) show homologous pairs of points, and (d) shows the original image: (a) between Frames 686 and 687 shown in Frame 686; (b) between Frames 687 and 688 shown in Frame 687; (c) between Frames 687 and 688 shown in Frame 688; (d) original image in Frame 689

Data Processing and Results

Image data processing mainly included camera lens distortion correction, Harris feature point extraction, image matching based on correlation coefficient, fundamental matrix fitting by RANSAC, and estimation of the trifocal tensor.

In camera lens distortions correction, the authors chose Frame 149 images for lens correction. The results of distortion correction are presented in Figs. 2(a–c), where Fig. 2(a) is the raw front view sequence image, Fig. 2(b) is the nondistortion image processed using the Brown model, and Fig. 2(c) is a difference image between the raw image and the nondistortion image. Fig. 2(c) shows that there is a small distortion in the center area of the image and a large distortion at the image edge and surrounding area that is the primary source of systematic error.

For the Harris feature point extraction, the authors chose the continuous four-frame nondistortion image sequence from Frame 686 to Frame 689 to extract feature points using the Harris corner detector operator. For parameters, the standard deviation (σ) of the Gaussian was 1, the search radium of the small shift was set to 3 pixels, and the number of corner points was 500. The extraction of feature points resulted in 3,024 points on Frame 686, 3,287 points on Frame 687, 3,348 points on Frame 688, and 2,541 points on Frame 689. The results are shown in the subplots in Figs. 3(a–d). The extracted feature points are shown in the corresponding images, respectively.

In the sequence image matching based on correlation coefficient, the window size for correlation matching was set as 11 × 11, the match radium was 5, the maximum search distance for matching was 50 × 50, and the value of the correlation coefficient was set as 0.99. The matching results in the sequence of four adjacent images are shown in subplots Figs. 4(a–c). The matched homologous pairs of points between Frames 686 and 687 are shown in Frame 686 in Fig. 4(a). The homologous pairs of points between Frames 687 and 688 are shown in Frame 687 in Fig. 4(b). The homologous pairs of points between Frames 688 and 689 are shown in Frame 688 in Fig. 4(c). The original image of Frame 689 is shown in Fig. 4(d).

In fitting the fundamental matrix of two views with the RANSAC algorithm, the authors chose the parameters $s = 8$, $p = 0.99$, $\varepsilon = 5\%$, and $t = 0.002$. The inlier matched results in the adjacent four-image sequence are shown in the subplots in Figs. 5(a–c). The result of inlier point fitting by RANSAC for Frames 686 and 687 is shown in Frame 686 in Fig. 5(a), that for Frames 687 and 688 is shown in Frame 687 in Fig. 5(b), and that for

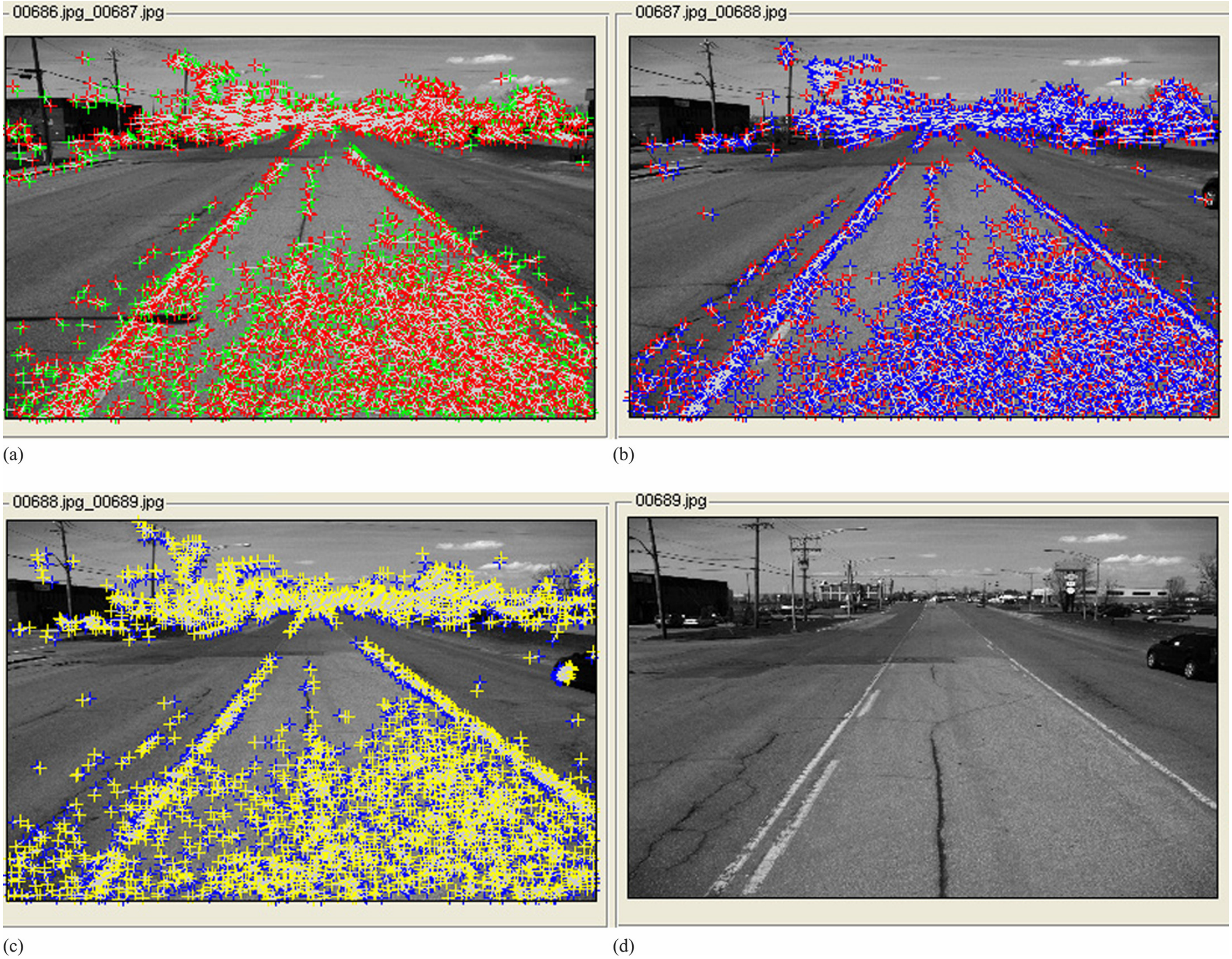


Fig. 5. Subplots (a) through (c) show inlier points fitted by RANSAC, and (d) shows the original image: (a) between Frames 686 and 687 shown in Frame 686; (b) between Frames 687 and 688 shown in Frame 687; (c) between frames 688 and 689 shown in frame 688; (d) original image in Frame 689

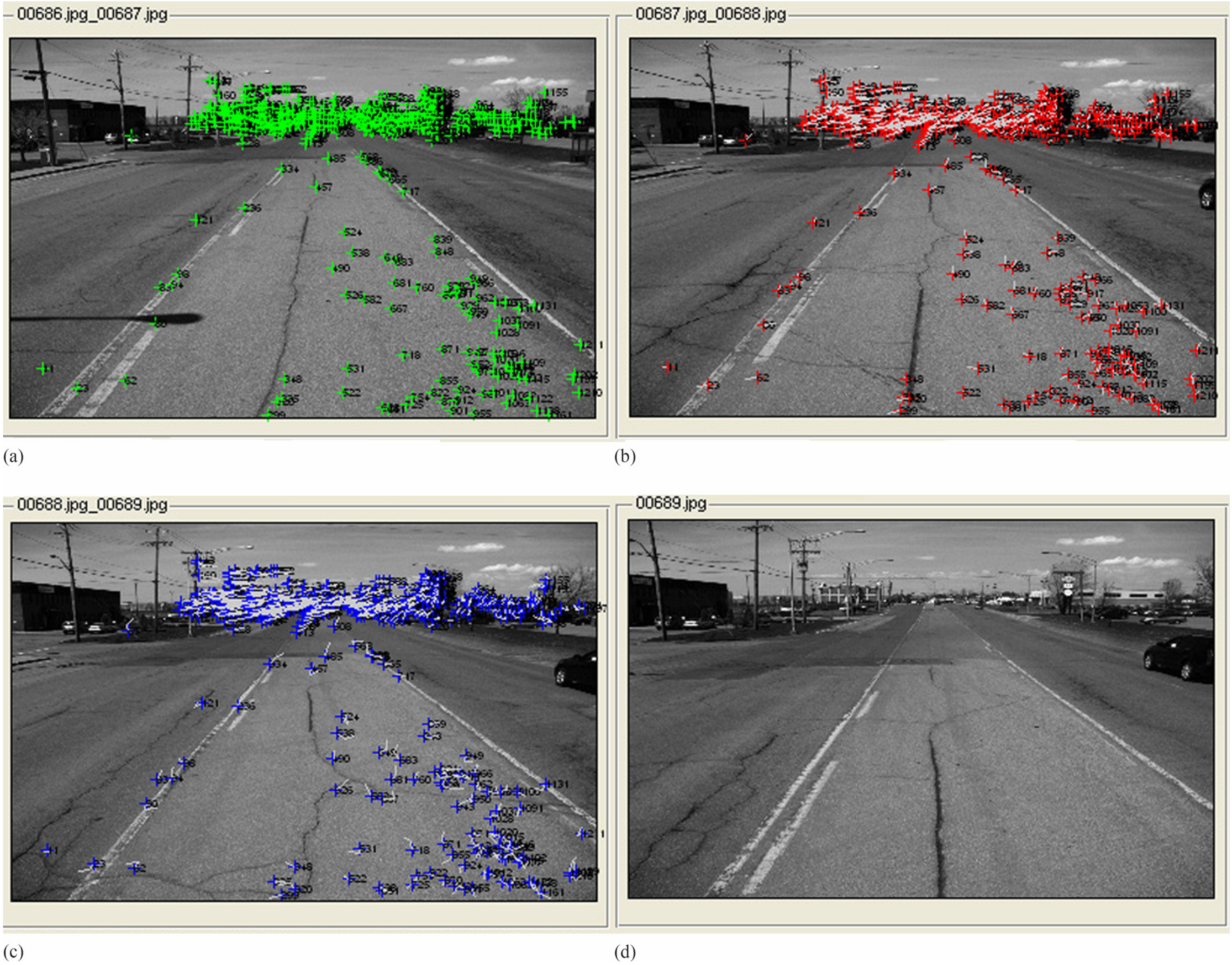


Fig. 6. Subplots (a) through (c) show triplet pairs of points, and (d) shows the original image: (a) between Frames 686 and 687 of three views shown in Frame 686; (b) between Frames 687 and 688 shown in Frame 687; (c) between Frames 688 and 689 shown in Frame 688; (d) original image in Frame 689

Frames 688 and 689 is shown in Frame 688 in Fig. 5(c). The original image of Frame 689 is shown in Fig. 5(d).

In computing the trifocal tensor, the triplet points were found from matched pairs of points between Frames 686 and 687 and Frames 687 and 688, and the triplet points were obtained from the matched pairs of points between Frames 687 and 688 and Frames 688 and 689. Then, the triplet points were estimated using the RANSAC algorithm, and the trifocal tensor was computed. The trifocal tensor $T_{686\_7\_8}$ in Frames 686, 687 and 688 was computed as follows; the result of the number of triplet points was 507 in those three images, respectively:

$$T_{686\_7\_8} = \left[\begin{array}{cccc} -242.796742537499 & -308.3119278271618 & -0.168649231801065 & \\ 83.98159742299697 & 0.63647443276779 & -0.0007727850778 & \\ 0.053243152720731 & 0.003388789008461 & 0.000001174601933 & \\ & -12.492666496867255 & 96.96361502455825 & -0.003774280617473 \\ & -359.74694098738854 & -233.03056978126986 & -0.170093451162632 \\ & 0.053243152720731 & 0.003388789008461 & 0.000002819043955 \\ & 13989.89991922595 & 19639.11457849911 & 1156195430323927 \\ & -4468.773116527736 & 1742.9776543351472 & 82.35843484120251 \\ & -350.0665863770545 & -304.8744643696752 & -0.117043925213603 \end{array}\right]$$

The results of the matched triplet points in the sequence of three adjacent images by trifocal tensor are shown in the subplots in Figs. 6(a–c). The triplet pairs of points between Frames 686 and 687 in the three images of Frames 686, 687, and 688 found by computing the trifocal tensor are shown in Frame 686 in Fig. 6(a); the triplet pairs of points between Frames 687 and 688 in the three images of Frames 687, 688, and 689 are shown in Frame 687 in Fig. 6(b); and the triplet pairs of points between Frames 688 and 689 in the three images of Frames 687, 688, and 689 are shown in Frame 688 in Fig. 6(c). The original image of Frame 689 is shown in Fig. 6(d).

Conclusions

This paper presents an approach to acquiring highly accurate matched points of sequence images from a MMS. Sequence images of motion with short baseline and overlapping are collected, and an image scale of front-view images is changed in accordance with the speed of the MMS. The approach applies the algorithms of computer vision to correct the camera lens distortion, to extract feature points using the Harris operator, and to produce many matched homologous points with higher matching effectiveness and many mistakenly matched points. The fundamental matrix of two views is computed by using the epipolar geometric constraint and RANSAC strategy to robustly estimate and effectively eliminate outliers. The accuracy and reliability of the matched points are thus improved. On the condition of error constraints based on minimizing algebra and homologous matched points of multiviews, the trifocal tensor is computed through the fundamental matrix and RANSAC algorithm. The high-precision public triplet points in three views are found. At the same time, the number of matched points in three views of sequence images is reduced to satisfy the trifocal tensor. The results of application of the method demonstrate that the proposed algorithm is very promising in terms of both accuracy and feasibility. In future work, the authors will solve the 3D point coordinates by bundle adjustment over three views using triplet pairs of points and EO parameters from GNSS/IMU data.

Acknowledgments

The authors gratefully acknowledge the use of sequence images supplied by Mankit Liu of CASW Instruments, Hong Kong, as a general agent of Applanix Landmark MMS, a Trimble Company, and thank Professor Yongshen Zhang of the Zhengzhou Institute of Surveying and Mapping, China, for instruction. The authors thank the China Scholarship Council (CSC) for financial aid, and the work was partially funded by the State Key Laboratory of Geo-information Engineering (No. SKLGIE2016-M-3-3). The authors thank the anonymous reviewers for their very helpful comments and suggestions on early drafts of this paper. The authors also thank the editors for their assistance.

References

Brown, D. C. (1971). “Close-range camera calibration.” *Photogramm. Eng.*, 37(8), 855–866.

Brown, D. C. (1976). “The bundle method—Progress and prospects.” *Int. Arch. Photogramm.*, 21(3), 1–33.

Brown, M., Szeliski, R., and Winder, S. (2005). “Multi-image matching using multi-scale oriented patches.” *Proc. of the 2005 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition*, IEEE, New York, 510–517.

Elaksher, A. F. (2008). “Multi-image matching using neural networks and photogrammetric conditions.” *Proc., Int. Archives of the Photogrammetry, Remote Sensing, and Spatial Information Sciences*, Vol. XXXVII, International Society for Photogrammetry and Remote Sensing, Hanover, Germany, 39–44.

Fishler, M. A., and Boles, R. C. (1981). “Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography.” *Commun. ACM*, 24(6), 381–395.

Forstner, W., and Gulch, E. (1987). “A fast operator for detection and precise location of distinct points, corners and centers of circular features.” *Proc., ISPRS Intercommission Workshop on Fast Processing of Photogrammetric Data*, International Society for Photogrammetry and Remote Sensing, Hanover, Germany, 281–305.

Gruen, A. W. (1985). “Adaptive least squares correlation: A powerful image matching technique.” *S. Afr. J. Photogramm., Remote Sens., Cartogr.*, 14(3), 175–187.

Harris, C., and Stephens, M. (1988). “A combined corner and edge detector.” *Proc., 4th Alvey Vision Conf.*, University of Sheffield Printing Office, Sheffield, U.K., 147–151.

Hartley, R., and Zisserman, A. (2000). *Multiple view geometry in computer vision*, Cambridge University Press, Cambridge, U.K.

Hartley, R., and Zisserman, A. (2003). *Multiple view geometry in computer vision*, 2nd Ed., Cambridge University Press, Cambridge, U.K.

Lepetit, V., Moreno-Noguer, F., and Fua, P. (2009). “EP*n*P: An accurate O (n) solution to the P*n*P problem.” *Int. J. Comput. Vision*, 81(2), 155–166.

Longuet-Higgins, H. C. (1981). “A computer algorithm for reconstructing a scene from two projections.” *Nature*, 293, 133–135.

Maas, H.-G. (1996). “Automatic DEM generation by multi-image feature based matching.” *Proc., Int. Archives of the Photogrammetry, Remote Sensing, and Spatial Information Sciences*, International Society for Photogrammetry and Remote Sensing, Hanover, Germany, 484–489.

Masry, S. E. (1981). “Digital mapping using entities—A new concept.” *Photogramm. Eng. Remote Sens.*, 47(11), 1561–1565.

Nister, D. (2004). “An efficient solution to the five-point relative pose problem.” *IEEE Trans. Pattern Anal. Mach. Intell.*, 26(6), 756–770.

Philip, J. (1996). “A non-iterative algorithm for determining all essential matrices corresponding to five point pairs.” *Photogramm. Rec.*, 15(88), 589–599.

Triggs, B., McLauchlan, P. F., Hartley, R. I., and Fitzgibbon, A. W. (2000). “Bundle adjustment—A modern synthesis.” *Proc., Int. Workshop on Vision Algorithms: Theory and Practice*, Springer-Verlag, London, 153–177.