A marker-less monocular vision point positioning method for industrial manual operation environments

Junhao Geng1 · Xinyang Zhao1 · Zhenxin Guo1 · Shangan Zhang2 · Jianjun Tang3 · Yiming Rong4 · Yongsheng Ma4,5

Received: 23 November 2021 / Accepted: 24 March 2022 / Published online: 5 April 2022
© The Author(s), under exclusive licence to Springer-Verlag London Ltd., part of Springer Nature 2022

Abstract
Vision-assisted technologies in industrial manual operation such as augmented reality (AR) are increasingly popular. They require high positioning accuracy and robustness to operate normally. However, narrow spaces, moving hands, or tools may occlude or obscure local visual features of operation environments and thus affect the positioning accuracy and robustness of operating position. The resultant misguidance may even cause misoperation of operators. This paper proposes a marker-less monocular vision point positioning method for vision-assisted manual operation in industrial environments. The proposed method can accurately and robustly locate the target point of operation using constraint minimization method even if the target area has no corresponding visual features in the case of occlusion and improper illumination. The proposed method has three phases: intersection generation, intersection optimization, and target point solving. In the intersection generation phase, a certain number of intersections of epipolar lines are generated as candidate target points using fundamental matrices. Here, the solving constraint is converted from point-to-line to point-to-points. In the intersection optimization phase, the intersections are optimized to two different sets through the iterative linear fitting and geometric mean absolute error methods. Here, the solving constraint is further converted from point-to-points to point-to-point sets. In the target point solving phase, the target point is solved as a constrained minimization problem based on the distribution constraint of the two intersection sets. Here, the solving constraint is finally converted from point-to-point sets to point-to-point, and the unique optimal solution is obtained as the target point. The experimental results show that this method has a better accuracy and robustness than the traditional homography matrix method for the practical industrial operation scenes.

Keywords Visual guidance · Industrial operation · Point positioning · Constrained minimization · Epipolar constrain

1 Instruction

With the rapid development of visual assistance technologies such as augmented reality (AR), visual-assisted industrial manual operation has been recognized as a promising technology in industrial environments such as assembly and maintenance phases of automotive, aerospace, and military equipment domains [1–4]. At present, marker-less monocular visual-assisted systems based on the natural features of operation environments has become a mainstream technology considering its practicability and low cost. Meanwhile, it requires minimum or zero setup efforts of end users and can be adapted to an unprepared environment [5–7]. However, there are still some challenges to overcome for more effective application [4]. In the case of occlusion and improper illumination, how to accurately and robustly locate the operation position with marker-less monocular images is a critical problem that has not been solved effectively.
One major task of industrial manual operations is to identify and locate the operation locations which usually are some specific spatial points, and then carry out the handling, aligning, joining, disjoining, adjusting, or other actions based on these locations [7–9]. The operator needs accurate and visual guiding information based on the locations to complete the operation correctly, especially for the small objects such as the holes of engine crankcast [10–13]. These manual operations with high precision and robustness requirements mostly occur in industrial environments. The narrow operation space and moving hands or handheld tools may cause full occlusion or improper illuminations around the target point. Therefore, the accuracy of monocular positioning methods based on visual features may be affected, and even wrong locations will be determined.

The operator needs to complete the work with bare hands or handheld tools because most of industrial operations are typically carried out on-site and heavily depend on manual maneuver [14]. Sometimes, the operator also carries out the tasks in some unprepared environments with uncertain illuminations. But the hands of operator or handheld tools may fully or partially occlude the target area, and the strong light may also obscure the target [5, 6, 15–17]. This state will result in that the target area loses the recognizable features used to retrieve or match the target point [18]. At the same time, the state without target features will last for a period of time along with the action of the operator. Furthermore, the camera pose may also change in the operation process. Accordingly, the target point will not be automatically identified and located because of no recognizable features in an unsteady image. Therefore, the corresponding visual guiding information could not be registered in the vision scenes based on the target location [19].

Lots of research fruits have been presented in the object identification and tracking in the case of occlusion or changeable illumination [20, 21] while few in accurate and robust methods of point positioning in the case of full and durable occlusion or obscuring, marker-less monocular cameras, and practical industrial environments. It may be partly due to that the common method, feature matching, will not be good effect in this case. In this situation, the homography matrix (HM) method could be used to approximately estimate the location of the target point while it may bring a non-ignorable registration error in complex environments, because HM method is more suitable for planar environments, but most of industrial environments are spatial [22]. Nevertheless, it is still possible to solve this problem with epipolar geometry method because it actually captures some information of scene structure [23]. However, the epipolar constraint of multiple view geometry in computer vision cannot handle this situation because it is a constraint of point-to-line rather than point-to-point [24, 25]. If the constraint of point-to-line can be converted to point-to-point, the problem of accurate and robust positioning may possibly be solved. Therefore, we propose a novel method to position the target point accurately and robustly through solving a constraint minimization problem with a unique solution.

In this paper, we address the challenge of accurate and robust marker-less monocular vision positioning in vision-assisted industrial operation scenes without target features. The main contributions of our work are as follows: (1) a better marker-less monocular vision point positioning method which can accurately and robustly locate the target point without around visual features and can be applied in practical industrial environments and (2) a novel constrained minimization solving algorithm for point correspondence in industrial computer vision which converts the constraint from point-to-line to point-to-point and get a unique solution. The experimental results show that this method has a better accuracy and robustness than the approximate estimate method based on HM and meets the requirements of visual guidance for practical industrial environments through being integrated in the visual guidance process such as AR guidance.

This paper is organized as follows. Section 2 introduces previous research works in this field. Section 3 gives the details of proposed method. Section 4 shows the experimental settings and results, assesses the performance, and discusses the results and limitations of the proposed method. Finally, conclusions and future work are presented in Sect. 5.

2 Related works

2.1 Point positioning in industrial operation environments

Many tasks of industrial operations need to be carried out based on a specific and precise location such as connector wiring and micro drill clamping [2, 7–9, 26, 27]. However, hands and handheld tools often fully or partially occlude the area of target location because most stages of operation activities are carried out manually by operators [2, 15, 28]. The illumination changes because of the narrow space or the reflection on metal surface may also obscure the target area. Accordingly, the visual guiding information such as symbols and cues could not be registered and displayed on the correct position because the target point may be positioned with non-ignorable errors [11, 28, 29]. Incorrect position guidance may lead serious operation quality problems, and the benefits of vision-assisted systems cannot be maximized [27]. Therefore, how to accurately and robustly locate the target point when the target area has no valid features is a critical issue for vision-assisted systems for industrial operations such as AR applications.
Losing target features is one of the most common problems of computer vision positioning. Losing target features is the state when the target is hidden or obscured partially or fully by other objects or light in the scene and severely affects the detection of target [19, 20]. It is a problem as well as a challenge to locate or track the target when its appearance or some key attributes are not available for the camera while the target is still present in the scene [19, 30]. Some researchers have tried various methods to solve this problem in computer vision scenes of industrial fields.

Zubizarreta et al. [15] introduced a matching scheme and template-based optimization using corresponding conics between model surface circles and image ellipses for 3D non-Lambertian object recognition in arbitrary environments. This method could detect the visible geometry features for the current pose of target object applying occlusion queries based on the z-buffer. But this method needs the offline training with CAD models and would fail if the heavy occlusion occurs. Gao et al. [31] used a locally supported Gaussian weight function and a bilateral filtering and outlier removal to handle cluttered scenes with partial occlusions for robust object recognition and matching. This method can identify the object in cluttered scenes with partial occlusions and build virtual-real object registrations via point cloud fusion. Wang et al. [16] used a tracking algorithm combining visual feature matching and point cloud alignment to achieve a good tracking performance under the partial occlusion condition. A reference point cloud model generated from its 3D model in a computer-aided design system is required for the establishment of benchmark coordinate system. Wang et al. [32] adapted a LINE-MOD method into a scale-invariant descriptor using depth information to handle occluding boundaries of the target object. This method needs the offline training with CAD models and is fast and robust for partial occlusion rather than full occlusion because the adopted template method would fail in the case of full occlusion. Huang et al. [33] used a monocular real-time robust feature tracking algorithm (MRRFT) to track a deep space target even it is partially occluded by a strong light spot. This method needs a point set to track a target object rather than a point, and the occluded points would not be tracked. Wang et al. [34] used a new image feature named as chain-of-lines feature (COLF) constructed by several directed line segments to register 3D objects in the camera view. This method can establish multiple correspondences between the 2D scene image and 3D model simultaneously even in a complex assembly environment with partial occlusion. This method also needs the offline training with CAD models and is only suitable for partial occlusion rather than full occlusion.

It can be seen that above methods are valuable for tracking a target object rather than positioning a target point. How to locate a target point in the demanding operation environments still needs to be studied. For the point positioning, Lima et al. [6] adopted a model-based tracking method to correctly identify points in tasks that involved tracking a rotating vehicle. However, the user needs to select each corresponding point from one valid key frame in the same sequence of the reference 2D points in order to correctly match these 3D points. In general, these methods are effective when the targets are partially occluded and are more suitable for a prepared environment because most of them need off-line training based on CAD models and also need a point cloud obtained by depth camera scanning. Actually, the marker-less monocular tracking is adopted extensively in industrial practice because it just requires minimum or zero setup efforts of end users and can be adapted to the unprepared environment [5, 7]. Therefore, how to precisely locate a target point rather than an object when the target area is fully occluded or obscured in an unprepared environment of industrial operation still needs to be further investigated. In other words, a method which can accurately find a single point correspondence between template image and target image without valid features of target area needs a further research.

### 2.2 Point correspondence under epipolar constraint

Positioning a target point in the target image based on a reference point of the template image is actually a point correspondence problem. The point correspondence can be considered as a part of the point set registration or image registration process in computer vision [35, 36]. The main purpose of the point set registration aims to find correspondences of points and to estimate the transformation between two or more point sets [37]. Image registration, also known as image fusion, matching, or warping, is the process of overlaying two (or more) images of the same area through geometrically aligning common features (or control points) identified in the images [38]. In practice, both of point set registration and image registration suffer from many challenges including occlusion or obscuring [35, 39]. The target point cannot be matched well in the case of occlusion or obscuring with common registration methods, for example, the feature matching–based methods and the iterative closest point algorithm, because the corresponding features of target point are insufficient or missing. Therefore, the location of target point can only be estimated and searched according to the reference point in the template image as well as the constraint of multiple view geometry.

The epipolar constraint is the only geometric constraint between two uncalibrated images of the same scene observed from two different camera viewpoints [40]. The epipolar constraint means a mapping between the point on the template image and the epipolar line on the target image. For each reference point in the template image, we
could limit the search for a corresponding target point in the target image to just an epipolar line (instead of naively searching the whole target image) [25]. However, the specific target point cannot be located without effective auxiliary information such as the depth because the epipolar constraint is not a one-to-one constraint [25]. In practice, the auxiliary information is unavailable for the uncalibrated images from unprepared environments, especially demanding industrial environments. On this condition, the homography matrix can be used to approximately estimate the target point through the perspective transform based on the homography relation [24, 41].

The mapping of pixel points, however, is accurate only in the case of pure camera rotation or planar scene [42]. The demanding industrial environments mostly are spatial rather than planar and the cameras also have translation as well as rotation which may severely affect the positioning accuracy based on the homography matrix. A large positioning error could lead to incorrect guiding information, especially in the case of multiple small targets, which may result in a wrong action. In order to locate the target point, Wang et al. proposed a dual-correlation method to determine the relationship between matching points in different images by a pair of different correlation transformations. This method uses two fundamental matrices (FMs) to generate two pairs of epipolar line and get the intersection of two lines as the target point [43]. Although this method can determine a single point correspondence between two uncalibrated images with fundamental matrix, it does not take into account the uncertainty of intersection. The intersection may be far away from the target point because of the uncertainty of FM.

In computer vision, the FM is the algebraic representation of the epipolar geometry that relates two images of a scene observed from two different viewpoints and describes the geometric relation of the two images [44, 45]. If the internal (intrinsic) parameters of camera are unknown, the pixel image coordinates should be used, and the matrix is known as the fundamental matrix [25]. In practice, the accuracy of FM is affected by the matching accuracy of control points and their depth variation [23, 44–46]. Although errors always exist, the target point is still distributed in the narrow region of the epipolar envelope with some fixed probability [23, 47, 48]. Stojanovic et al. [47] used Monte Carlo simulation to obtain a set of different FMs and computed an envelope of epipolar lines. The potential target point has a maximal probability in this area. This method needs to select specific eight correspondences for computing the different FMs. However, the specific correspondences cannot reduce the tendency of location error because of the uncertainty of the FM. At the same time, the target point still cannot be uniquely obtained because this is a constraint of point-to-area rather than point-to-point.

Generally, existed methods have proved that the FM is dependent of the scene structure and pointed out the most probable distribution area of the target point. Our investigation also shows that there is still no better method than the approximate estimation based on the homography matrix to uniquely position the target point in the case of uncalibrated images, spatial scene, complete and durable target occlusion/obscurring, and monocular marker-less vision. However, it can be seen that the search space of the target point has been reduced from a line to a narrow region based on the uncertainty of the FM [23, 43, 47]. If the geometric constraint can be further reduced into a constraint of point-to-point, it would be possible to accurately position the most likely target point uniquely.

2.3 Summary

Most of related researches in the field of practical industrial environments focus on the positioning of an object under partial occlusion rather than a point under complete occlusion or obscuring and have achieved good results. Generally, these methods need offline training based on 3D models in advance and the depth information of points in runtime. Therefore, these methods are not suitable for the target point positioning in the case of unprepared environments, marker-less monocular vision, uncalibrated cameras, and full occlusion or obscuring. If the target point positioning problem is abstracted as a geometric constraint problem of computer vision, the epipolar constraint based on the FM is the main theoretical basis related to this problem. At present, the related literatures which have promoted the understanding of epipolar geometry mainly focus on the registration of point set or image rather than a single point correspondence. How to uniquely and accurately position the target point in the demanding industrial environments is still a problem that needs further study. However, these researches simultaneously inspire us that the uncertainty of the fundamental matrix is conducive to determine the distribution and probability of the target point. Hence, we proposed a novel method that can uniquely and accurately position the target point without its feature information. This method has a better accuracy than the approximate estimate method based on the homography matrix.

3 Proposed method

3.1 Overview

In this paper, we present a marker-less monocular vision point positioning method with higher accuracy and robustness through directly positioning the target point based on the location of designated reference point. This method is
suitable for the vision-assisted scenes of industrial manual operations in the case of spatial environments, full and durable target occlusion or obscuring, and uncalibrated monocular cameras. The main phases of the general positioning process can be summarized as follows: image preprocessing, point feature extraction, point feature matching, and visual content registration [35, 38]. If the target area is completely occluded or obscured for some time whilst the position of camera changes with the operator’s action, the steps of point feature extraction and matching based on the natural features will not work properly. Our method can find the position of the target point through replacing the phases of point feature extraction and matching in the existing methods. The process is presented in Fig. 1.

The core idea of this method is to convert the constraint of point-to-line with uncertainty to the constraint of point-to-point with unique solution. In other words, this method can reduce the solution number of positioning problem from many uncertain solutions to a constrained minimization solution. Hence, the target point could be located uniquely and accurately. The workflow is separated in five phases: image preprocessing, intersection generation, intersection optimization, target point solving, and visual content registration.

Image preprocessing is the first phase which accepts the input images, detects and matches their feature points, and optimizes the matched feature points. If the two images are matched, the optimized matched feature points will be output to the next phase. But the feature points in the target area will not be matched because of the occlusion or obscuring of the target area even they are detected.

The second phase, intersection generation, uses the matched feature points to generate a certain number of epipolar line pairs with adding Gaussian noise to the matched feature points, and then calculates the intersections of epipolar line pairs as the candidate solutions of target point.

The third phase, intersection optimization, detects two optimized intersection sets based on the methods of iterative linear fitting (ILF) and geometric mean absolute error (GMAE), respectively. These two sets have specific constraint relation and are used to construct a constrained minimization problem and solve the optimal solution.

The fourth phase, target point solving, transforms the constraint relation between the function representations of two sets to a constrained minimization problem, then constructs the objective function and constrained function to solve the optimal solution based on the Lagrange multiplier method. The optimal solution just is the desired target point.

The last phase, visual content registration, transforms the coordinates of the target point to the screen coordinates and superimposes the visual contents based on the screen coordinates of the target point. Finally, the visual scene with right guiding contents can be shown to the operator and the point positioning process ends.

The first and last phases are the essential procedures and can be implemented through the existing general technologies. They provide the inputs and display the results. So, our method will focus on the other three intermediate phases.

The problem solved by our method is stated as follows: given two sets of matched featured points $PS_R$ and $PS_T$ from the template image and the target image respectively, find the location of the target point $P_T(x_T, y_T)$ in the target image based on the location of the reference point $P_R(x_R, y_R)$ in the template image. The taking angle and position of two images are different. The poses and intrinsics of monocular camera are uncalibrated. No corresponding feature points can be matched around the target point, and no markers are on the two images. Our method needs to use above three phases to position the pixel coordinates of the target point according to these limited conditions.

### 3.2 Intersection generation

The phase of intersection generation accepts the matched feature points generated from the image preprocessing phase as input parameters. At the image preprocessing phase, SURF, SIFT, or other common methods can be adopted to detect the feature points. And RANSAC or other general methods can be used to match feature points [35, 36, 49]. After image preprocessing, it should be confirmed that the template image and target image are matched, and the matched points are obtained as shown in Fig. 2. This is the premise of positioning the target point. The target point cannot be located based on two unmatched images.

On account of the matched featured points, a fundamental matrix $F$ can be obtained through RANSAC or other methods [24]. The fundamental matrix is the algebraic representation of epipolar geometry and is used to get a corresponding epipolar line in the target image for a reference point in the template image. For each reference point $P$ in the template image, there exists a corresponding epipolar line $L$ in the target image. Any target point $P'$ in the target image matching the point $P$ must lie on the epipolar line $L$. But the exact location of the target point $P'$ on the epipolar line is unknown because the epipolar line is the projection of the ray of light $R$ from the reference point through the camera center $C_R$ rather than a point as shown in Fig. 3. If this constraint of point-to-line can be transformed to the constraint of point-to-point, the location of the target point would be determined.

A straightforward idea is to find another epipolar line in the target image for the reference point. Hence, the intersection of two epipolar lines will be the target point because it should be on both lines at the same time [43]. However, a single intersection is unreliable because of the uncertainty of fundamental matrix itself [23, 46]. The most likely target
points are located in a hyperbolic region as shown in Fig. 4. So, a certain number of likely target points rather than one point should be obtained in order to retrieve the correct target point. Our method to obtain intersections is divided into three steps: generation of fundamental matrix pairs, generation of epipolar line pairs, and generation of intersections.

In the step of generation of fundamental matrix pairs, a certain number pairs of fundamental matrices are generated...
through adding zero-mean Gaussian noise to the matched feature points. If there are \( n \) feature points in the set of matched feature points \( PS_R \) and \( PS_T \) respectively, and the feature point \( P_i \) in \( PS_R \) and the matched feature point \( P_{i}^{'} \) in \( PS_T \) are the correspondence points \( P_i \rightarrow P_{i}^{'} \), a fundamental matrix \( F \) can be obtained through the 8-point method \( [24, 25] \). Here, we use the RANSAC with normalized eight-point method to calculate all correspondence points instead of 8 specific correspondences for obtaining \( F \) as shown in Eq. (1) because all correspondence points can maximize the overall robustness of \( F \):

\[
8\text{PointsRANSAC}(PS_R = \{P_1, \ldots, P_n\}, PS_T = \{P'_1, \ldots, P'_n\}) \Rightarrow F
\]  

(1)

If a zero-mean Gaussian noise with a standard deviation \( \sigma_x = \sigma_y \) is added to each feature point \( P(x, y) \) and \( P'(x', y') \) to get the noised feature points \( NP(x + N(0, \sigma_x), y + N(0, \sigma_y)) \) and \( NP'(x + N(0, \sigma_x), y + N(0, \sigma_y)) \), a noised fundamental matrix \( NF \) can be obtained:

\[
8\text{PointsRANSAC}(PS_R = \{NP_1, \ldots, NP_n\}, PS_T = \{NP'_1, \ldots, NP'_n\}) \Rightarrow NF
\]  

(2)

However, another \( NF \) should be obtained because two \( NF \)'s are needed to generate two epipolar lines for getting their intersections. In order to get another \( NF \) and make sure that two epipolar lines have an intersection, all feature point in \( PS_R \) and \( PS_T \) are reversed, and a new corresponding fundamental matrix \( NF' \) can be obtained:

\[
8\text{PointsRANSAC}(PS_R = \{NP_n, \ldots, NP_1\}, PS_T = \{NP'_n, \ldots, NP'_1\}) \Rightarrow NF'
\]  

(3)

So, a pair set of fundamental matrices \( PFS \) is obtained through \( m \) times iteration of adding Gaussian noise:

\[
PFS = \{(NF_1, NF'_1), \ldots, (NF_m, NF'_m)\}
\]  

(4)
In the step of generation of epipolar line pairs, \( m \) epipolar line pairs can be obtained based on the fundamental matrix pairs. A fundamental matrix is a \( 3 \times 3 \) matrix with rank 2. An epipolar line \( Ax + By + C = 0 \) can be expressed as \( L = [A, B, C]^T \). The reference point \( P_R(x_R, y_R) \) can be further expressed as \( P_R(x_R, y_R, 1) \) because it is a 2D point. The epipolar line \( L \) and its correspondence \( L' \) can be obtained through epipolar geometry:

\[
\begin{align*}
L &= NF \times P_R^T \\
L' &= NF' \times P_R^T
\end{align*}
\]  \( (5) \)

When all pairs of fundamental matrices are calculated, the pair set of epipolar lines \( PLS \) is obtained:

\[
PLS = \{(L_1, L'_1), \ldots, (L_m, L'_m)\}
\]  \( (6) \)

In the step of generation of epipolar line intersections, an intersection \( I(x, y, 1) \) of one pair of epipolar lines can be easily obtained through linear equations:

\[
\begin{bmatrix}
A & A' \\
B & B' \\
C & C'
\end{bmatrix}
\begin{bmatrix}
x \\
y \\
1
\end{bmatrix} = 0 \Rightarrow I = [x, y, 1]^T
\]  \( (7) \)

Finally, the set of intersections \( IS \) can be obtained:

\[
IS = \{I_1, \ldots, I_m\}
\]  \( (8) \)

Figure 5 shows an instance of noised feature points, corresponding epipolar line pairs and their intersections.

### 3.3 Intersection optimization

Although, theoretically, the intersections should be distributed in an ellipse or parabola area, they are always located in an hyperbola region named epipolar band in practice [23]. The target point would be located in the densest area of intersections where the two sections of hyperbolas are closest to each other [50]. Moreover, another theoretical distribution area should be a line because the epipolar constrain is the constraint of point-to-line. In other words, the target point should be located in the densest area and on the line at the same time. Therefore, the intersections should be optimized from two perspectives of linear fitting and point density.

An iterative linear fitting (ILF) method considering the distance between an intersection and the fitted line is proposed to optimize intersections from the perspective of linear fitting. The main idea of ILF is to fit a line with intersections and then remove outliers with a longer distance than a given distance. Next, the fitting and removing process is iterated, and the given distance is also gradually reduced until there are no outliers when the given distance is reduced to \( D_{\text{min}} \). The purpose of the distance iteration is to fit the most likely epipolar line as much as possible. \( Q \) denotes the total number of iterations, \( \Delta D \) denotes the distance increment per iteration. After \( i \)th iteration, the optimized intersection set is as follows:

\[
\begin{align*}
\text{LinearFit}(IS_{i-1}) & \Rightarrow L_i \\
IS_i &= \{I_1, \ldots, I_i\} | \text{Distance}(\forall I, L_i) < (D_{\text{min}} + (Q - i) \times \Delta D)
\end{align*}
\]  \( (9) \)

After iteration, \( LIS \), the final intersection set based on linear fitting optimization, is obtained. \( k \) intersections are assumed in \( LIS \):

\[
LIS = \{I_1, \ldots, I_k\} | \text{Distance}(\forall I, L_Q) < D_{\text{min}}
\]  \( (10) \)

In order to get the optimized intersection set from the perspective of point density, the geometric mean absolute error (GMAE) method is used to retrieve the densest points based on the given confidence interval \( \gamma \). \( E \), the GMAE of all \( I(x, y) \) in \( IS = \{I_1, \ldots, I_m\} \), can be calculated as follows:

![Feature points with N(0,1) noise](image1)

![One pair of epipolar lines](image2)

![1000 pairs of epipolar lines](image3)

![1000 intersections of epipolar line pairs](image4)

**Fig. 5** Noised feature points, epipolar line pairs and their intersections. a Feature points with \( N(0, 1) \) noise. b One pair of epipolar lines. c 1000 pairs of epipolar lines. d 1000 intersections of epipolar line pairs
\[ E = \sqrt{\left(\frac{1}{m} \sum_{i=1}^{m} |x_i - \bar{x}|^2 + \left(\frac{1}{m} \sum_{i=1}^{m} |y_i - \bar{y}|^2\right)^2 \right) } \]  
(11)

If the error of \( \forall i \) and \( \bar{I} \) is less than \( rE \), \( I \) will be remained as an inlier. Finally, \( DIS \), the optimized intersection set based on point density, is obtained. \( w \) intersections are assumed in \( DIS \):

\[ DIS = \left\{ I_1, \ldots, I_w | |\forall i - \bar{I}| < yE \right\} \]  
(12)

Figure 6 shows two optimized intersection sets from two perspectives of linear fitting and point density.

### 3.4 Target point solving

Now, two sets of optimized intersections have been obtained. Although the constraint of point-to-line has been transformed to the constraint of point-to-points, this is still a one-to-many problem. This problem should be reduced further to a one-to-one problem in order to get a unique target point. This problem can be transformed to a constrained minimization problem because the target point is distributed in a hyperbolic region which is constrained by the condition that the target point should also be on the fitted epipolar line.

According to the above analysis, we construct Lagrange functions to solve this constrained minimization problem with Lagrange multiplier. The optimal solution of this problem will be the target point. The objective function, \( f(x, y) \), is a minimal Mahalanobia Distance function which describes the distance between the target point and \( DIS \):

\[ f(x, y) = (I - \bar{I})^T \sum_+^I (I - \bar{I}) \]  
(13)

\( \sum_+^I \) denotes the SVD of inverse covariance matrix of \( DIS \) and is still a symmetric matrix. \( \bar{I} \) is the mean of \( DIS \):

\[ \sum_+^I = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix} \]  
(14)

So, \( f(X, Y) \) can be expanded as follows:

\[ f(x, y) = \begin{bmatrix} x - \bar{x} \\ y - \bar{y} \end{bmatrix}^T \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix} \begin{bmatrix} x - \bar{x} \\ y - \bar{y} \end{bmatrix} = \sigma_{11}x^2 + (\sigma_{12} + \sigma_{21})xy + \sigma_{22}y^2 \]  
(15)

The constraint function, \( g(X, Y) \), is a line function which is fitted with \( LIS \). The fitted line is denoted as \( FL = [A, B, C]^T \):

\[ g(x, y) = AX + BY + C \]  
(16)

The constrained minimization problem can be described as follows:

\[ \begin{cases} \min f(x, y) \\ g(x, y) = 0 \end{cases} \]  
(17)

The optimal solution can be found when the gradients of the two functions are parallel at the target point. The unknown constant multiplier \( \lambda \) is necessary because the magnitudes of the two gradients may be different:

\[ \nabla f(x, y) = \lambda \nabla g(x, y) \]  
(18)

After expanding Eq. 18, the following equations are determined:

\[ \begin{cases} \frac{\partial f(x, y)}{\partial x} = \lambda \frac{\partial g(x, y)}{\partial x} \\ \frac{\partial f(x, y)}{\partial y} = \lambda \frac{\partial g(x, y)}{\partial y} \\ g(x, y) = 0 \end{cases} \]  
(19)

After computing its derivatives, the constrained minimization problem can be reduced to solving a set of three linear equations as Eq. 20:

\[ \begin{cases} 2\sigma_{11} + \sigma_{12} + \sigma_{21} \\ 2\sigma_{22} + \sigma_{12} + \sigma_{21} \\ A & B \end{cases} \begin{bmatrix} x \\ 1 \end{bmatrix} = \begin{bmatrix} \lambda A \\ \lambda B \\ -C \end{bmatrix} \]  
(20)

This set of equations can be easily solved, and the solution vector \( (x, y, \lambda)^T \) is obtained. The target point \( P_T(x_T, y_T) = (x, y) + \bar{I} = (x + \bar{x}, y + \bar{y}) \) can be identified uniquely. Figure 7 shows the illustrative curves of the
objective function and constraint function, and the optimal solution which is the target point.

At last, visual contents can be superimposed on the visual scene based on the coordinates of target point. After the coordinates of target point are converted to screen coordinates, visual contents can be superimposed on the screen coordinates or other locations based on these coordinates as shown in Fig. 8. After that, the point positioning process ends and the superimposed visual contents with accurate and correct position will be shown for the operator.

4 Experiment and discussion

4.1 Experiment setup

There are no other direct methods to position a single point in the case of marker-less monocular image pairs and full or durative target occlusion or obscuring. The perspective transform method based on homography matrix can be used to approximately estimate the location of target point, but it is more suitable for planar environments. In fact, most industrial environments are not planar. Therefore, we designed an experiment which comes from practical industrial application and includes two kinds of typical industrial operation environments to evaluate the accuracy and stability of our proposed method over the homography matrix method.

Considering the relevance of the experiment and a better comparative analysis, we selected two industrial operation environments including aeroengine maintenance and CNC machine tool control. They are typical industrial manual operations. They are in typical industrial environments with occlusion or obscuration. The target area of aeroengine is occluded by operator’s hands, and the target area of CNC machine tool’s control panel is obscured by strong light. At the same time, these two environments
have different spatiality. The aeroengine is stereo, and the control panel is planar-like. Figure 9 shows the template images of two environments and one of their corresponding target images. Both methods use uncalibrated and marker-less monocular images to locate the target points.

In order to analyze the positioning accuracy and robustness in different positions and angles, we manually marked the target point on the fiducial image. The fiducial image was taken at the same taking position as the target image and without occlusion and obscuring. In each environment, 10 groups of images were taken at the different positions which have different translations and rotations relative to the taking position of the template image. These positions are evenly distributed in the range of the operator’s arms and the field of view angle. We set the left side of the reference camera as negative and the right side as positive. Figure 10 shows the layout setup of taking image for different groups. Table 1 shows these settings in detail. Table 2 shows the thumbnails of experimental images.

Our experiment was run on a ThinkPad computer with Windows 10 system, 2.8 GHz CPU, 16 GB memory, and integrated Intel HD 520 graphics card. The image resolution is 800 by 600 pixels. OpenCV 4.5.0 is used as development platform, and C++ is the used programming language.

### 4.2 Experiment results and discussion

We conducted the experiment based on the above settings. In order to evaluate the accuracy and robustness of the proposed method, and observe the influence of the number of intersections on them, we generated different numbers of intersections from 100 to 1000 to repeat the experiment. We used a common set of parameters. A zero-mean Gaussian noise with a standard deviation $\sigma = 1$, that is $N(0, 1)$, was added to each feature point. $D_{\text{min}} = 3$ and $\gamma = 2$ are set. All results’ errors of two environments are shown in Figs. 11, 12.

We calculated the experimental results, got the maximum $\max$, minimum $\min$, mean value $\mu$, standard deviation $\sigma$, and coefficient of variation $cv = \frac{\sigma}{\mu}$, and used them to evaluate the positioning accuracy and robustness of our method. Table 3 shows the computed results.

It can be seen that the accuracy of our method is better than the HM method for both environments according to $\mu$. From the perspective of the actual impact of errors, the difference between the minimum errors of the two environments is no more than two pixels, and so, there is no obvious impact on the visual guidance. However, the maximum error of the HM method for the aeroengine maintenance is too large (> 11 pixels) to be ignored, and this may cause incorrect operation. The stability of our method is also better than

---

**Fig. 9** Template images and one of their corresponding target images. (a) Aeroengine maintenance. (b) CNC machine tool control
Table 1  Experiment settings in detail

| Scene     | Spatiality | Type of occlusion/obscuring | Camera distance (cm) | Group name | Group ID | Rotation (degree) | Translation (cm) | Reference point/fiducial point |
|-----------|------------|-----------------------------|----------------------|------------|----------|-------------------|-------------------|-------------------------------|
| Aeroengine maintenance | Stereo      | Hands                        | 41                   | Template image | TE       | 0                 | 0                 | (454,250)                     |
|           |            |                              |                      | Target image L1 | L1       | -3                | -4.3              | (408,250)                     |
|           |            |                              |                      | Target image L2 | L2       | -6                | -8.6              | (359,250)                     |
|           |            |                              |                      | Target image L3 | L3       | -9                | -12.8             | (341,250)                     |
|           |            |                              |                      | Target image L4 | L4       | -12               | -17.0             | (324,250)                     |
|           |            |                              |                      | Target image L5 | L5       | -15               | -21.2             | (339,246)                     |
|           |            |                              |                      | Target image R1 | R1       | 3                 | 4.3                | (436,250)                     |
|           |            |                              |                      | Target image R2 | R2       | 6                 | 8.6                | (451,248)                     |
|           |            |                              |                      | Target image R3 | R3       | 9                 | 12.8               | (460,248)                     |
|           |            |                              |                      | Target image R4 | R4       | 12                | 17.0               | (479,248)                     |
|           |            |                              |                      | Target image R5 | R5       | 15                | 21.2               | (468,245)                     |
| CNC control | Planar-like | Strong light                 | 35                   | Template image | TE       | 0                 | 0                 | (408,335)                     |
|           |            |                              |                      | Target image L1 | L1       | -4.5              | -5.5              | (383,335)                     |
|           |            |                              |                      | Target image L2 | L2       | -9                | -11.0             | (369,336)                     |
|           |            |                              |                      | Target image L3 | L3       | -13.5             | -16.3             | (340,335)                     |
|           |            |                              |                      | Target image L4 | L4       | -18               | -21.6             | (328,335)                     |
|           |            |                              |                      | Target image L5 | L5       | -22.5             | -26.8             | (344,337)                     |
|           |            |                              |                      | Target image R1 | R1       | 4.5               | 5.5               | (430,336)                     |
|           |            |                              |                      | Target image R2 | R2       | 9                 | 11.0              | (459,336)                     |
|           |            |                              |                      | Target image R3 | R3       | 13.5              | 16.3              | (479,336)                     |
|           |            |                              |                      | Target image R4 | R4       | 18                | 21.6              | (491,336)                     |
|           |            |                              |                      | Target image R5 | R5       | 22.5              | 26.8              | (515,337)                     |
the HM method for both environments according to $cv$. This indicates that the error range of our method is smaller than the HM method even in the planar-like scene.

For the aeroengine maintenance, the error’s mean of proposed method, $\mu = 2.33$ pixels, is almost half that of the HM method. 2.33 pixels in the target image ($800 \times 600$ pixels) can be roughly converted to 1.6 mm in this corresponding physical scene ($50 \times 37.5$ cm). This average error can well meet the requirements of industrial operation guidance [31, 32]. For the CNC control, the mean error, $\mu = 0.89$, is more accurate. Actually, both methods have almost the same registration effect under this level of error.

The results also show that the single error of our method varies in different groups and with different intersections although the overall error is stable and accurate. The cause is that the optimal solution is obtained based on the statistical data. The single error of the HM method is the same with different intersections in the same group but varies in different groups. The cause is that the HM method runs based on the determined matched points of the image pair but not the statistical data.

We also observed that the errors of both methods have a similar changing trend. That means the quality of the matched points have the similar impact on them. If the
quality of matched points is not good, especially in the case of uneven distribution around the target point, the robustness of both methods will be affected. The results also show that the position of camera has no evident or direct impact on both methods. However, it seems that the spatiality of scene has impact on the accuracy of both methods. The closer the scene is to the plane, the smaller the error is. The fundamental matrix actually captures some information of scene structure; in other words, the fundamental matrix is dependent of the scene structure [23, 44–46]. Obviously, the proposed method is much less affected by the scene structure. This is mainly due to the application of statistical data and the constraint solving. Further research should be undertaken to investigate the inner influences of scene structure on the proposed method in order to get more accurate results.

In terms of efficiency, the proposed method is more time-consuming than HM method, because there is more computing to do. In general, the running time of HM method is less than 1 ms. Figure 13 shows the running time of the proposed method for the CNC control environment. The data show that running time increases with the number of intersections. However, we also could observe the number of intersections has almost no impact on the accuracy of the proposed method from Fig. 12. Selecting 100 intersections for calculation will take less than 10 ms. Although this runtime value is much higher than the runtime of HM method, this is acceptable for the real-time positioning and visual guidance in the case of a refresh rate of less than 60 frames. What is interesting about the data in Figs. 12, 13 is that the accuracy error and the calculation time have a certain positive correlation. We believe it is because the re-projection threshold of FM by RANSAC method needs more time to calculate when the error is large.

But some limitations of this method still exist. First, it needs a certain running time and mainly because of the generation of hundreds of fundamental matrices. Therefore, it is not suitable for the visual-aided applications with highly real-time requirements or a refresh rate of more than 60 frames. However, considering that the refresh rate of most visual-aided application is 30 or no more than 60, this method still has wide adaptability. Second, the robustness of this method may be influenced by the scene structure, i.e., the spatiality of operation environments. Excessively uneven spatiality of environment may lead to a greater positioning error. Actually, the HM method has the same problem, and its accuracy and robustness are worse than our method in this situation. In other words, this method has certain adaptability to the spatial environments. At last, this method is not suitable for the visual guidance which needs the same error for each positioning in the same environment. The registration error may change slightly because of adding Gaussian noise in calculating process. However, the accuracy still meets the requirements.

Overall, these results indicate that the proposed method has a better accuracy and robustness than the perspective transform method based on the homography matrix. Its accuracy and robustness meets the requirements of visual guidance for industrial operations. This method is more

Table 3  Experimental computed results

| Environment         | Method          | min  | max  | $\mu$  | $\sigma$ | $cv$ |
|---------------------|-----------------|------|------|--------|---------|------|
| Aeroengine maintenance | Proposed method | 0.12 | 4.96 | 2.33   | 1.10    | 0.47 |
|                     | HM method       | 1.8  | 11.37| 4.32   | 2.83    | 0.65 |
| CNC control         | Proposed method | 0.17 | 1.67 | 0.89   | 0.49    | 0.55 |
|                     | HM method       | 0.11 | 3.15 | 1.07   | 0.82    | 0.77 |

Fig. 13  Running time of the proposed method
suitable for the industrial operation scenes with higher positioning requirements, full and durative target occlusion or obscuring, marker-less and even spatial environments, and uncalibrated monocular cameras. It is also applicable to the general environments where the HM method is applicable. Only in terms of computational efficiency, if higher accuracy or higher real-time operation is not necessary, HM method can be selected a preferred way.

5 Conclusions and future work

Vision-assisted industrial operation has been recognized as a promising technology used in most of life-cycle phases and different fields. Accurate and robust point positioning in the case of no target features and uncalibrated monocular cameras still is a critical problem that has not been solved effectively. This study set out to design a marker-less monocular vision point positioning method for vision-assisted industrial operation environments in order to provide more accurate and robust visual guidance.

This study has shown that an accurate and robust point positioning problem can be solved by converting the constraint of point-to-line with uncertainty to the constraint of point-to-point with unique solution. This work contributes to existing knowledge of computer vision point positioning by providing a novel method with higher accuracy and robustness for practical industrial environments. It also contributes a novel constrained minimization solving algorithm for point correspondence in computer vision. The major limitations of this study are slightly larger amount of computation and influence of scene structure. Notwithstanding these limitations, the study can meet the practical industrial requirements of point positioning. Although this method can also be used in general environments, it is more suitable for the more demanding industrial scenes with higher positioning requirements, full and durative target occlusion or obscuring, marker-less spatial environments, and uncalibrated monocular cameras.

However, several questions still remain to be answered. One of them is to optimize the generation algorithm of fundamental matrices which is the main part of time consuming of this method. The other is to evaluate the method in more complex scene structures for more kinds of harsh industrial operations and further identify the mechanism how the scene structure affects the fundamental matrix as well as the position error.

Authors’ contribution All the authors contributed to the study conception and design. Material preparation, data collection, and analysis were performed by Junhao Geng, Xinyang Zhao, Zhenxin Guo, Shangan Zhang, Jianjun Tang, and Yiming Rong. The first draft of the manuscript was written by Junhao Geng and Yongsheng Ma, and all the authors commented on previous versions of the manuscript. All the authors read and approved the final manuscript.

Funding This work was supported by the National Natural Science Foundation of China under Grant 52175484 and was largely conducted when the first author was attached at the University of Alberta. The corresponding author, Prof. Yongsheng Ma, had been supported by Canada NSERC Discovery grant; he has moved from University of Canada to Southern University of Science Technology in Shenzhen, China, since July, 2021.

Data availability Data sharing not applicable to this article as no datasets were generated or analyzed during the current study.

Code availability Not applicable.

Declarations

Ethics approval Not applicable.

Consent to participate Not applicable.

Consent for publication Not applicable.

Competing interests The authors declare no competing interests.

References

1. Lopik KV, Sinclair M, Sharpe R, Conway P, West A (2020) Developing augmented reality capabilities for industry 4.0 enterprises: lessons learnt from a content authoring case study. Comput Indus 117(2020): 103208. https://doi.org/10.1016/j.compind.2020.103208

2. Bottani E, Vignali G (2019) Augmented reality technology in the manufacturing industry: a review of the last decade. IIE Transactions 51(3):284–310. https://doi.org/10.1080/24725854.2018.1493244

3. Palmarini R, Erkoyuncu JA, Roy R, Torabmostaedi H (2018) A systematic review of augmented reality applications in maintenance. Robot Comput Integr Manuf 49(2018):215–228. https://doi.org/10.1016/j.rcim.2017.06.002

4. Wang X, Ong SK, Nee AYC (2016) A comprehensive survey of augmented reality assembly research. Adv Manuf 4(1):1–22. https://doi.org/10.1007/s40436-015-0131-4

5. Nee AYC, Ong SK, Chrysosoulouris G, Mourtzis D (2012) Augmented reality applications in design and manufacturing. CIRP Ann 61(2):657–679. https://doi.org/10.1016/j.cirp.2012.05.010

6. Lima JP, Roberto R, Simões F, Almeida M, Figueiredo L, Teixeira JM, Teichrieb V (2017) Markerless tracking system for augmented reality in the automotive industry. Expert Syst Appl 82(2017):100–114. https://doi.org/10.1016/j.eswa.2017.03.060

7. Siew CY, One SK, Nee AYC (2019) A practical augmented reality-assisted maintenance system framework for adaptive user support. Robot Comput Integr Manuf 59(2019):115–129. https://doi.org/10.1016/j.rcim.2019.03.010

8. Radkowski R, Herrema J, Oliver J (2015) Augmented reality-based manual assembly support with visual features for different degrees of difficulty. Int J Human-Comput Int 31(5):337–349. https://doi.org/10.1080/10447318.2014.994194

9. Henderson SJ, Feiner S (2011) Exploring the benefits of augmented reality documentation for maintenance and repair. IEEE
Geng J, Tian X, Bai M, Jia X, Liu X (2014) A design method for three-dimensional maintenance, repair and overhaul job card of complex products. Comput Ind 65(1):200–209. https://doi.org/10.1016/j.compind.2013.08.008

Gattullo M, Scurati GW, Fiorentino M, Uva AE, Ferrise F, Bordegioni M (2019) Towards augmented reality manuals for industry 4.0: a methodology. Robot Comput Integr Manuf 56 (2019):276–286. https://doi.org/10.1016/j.rcim.2018.10.001

Tatić D, Tesić B (2017) The application of augmented reality technologies for the improvement of occupational safety in an industrial environment. Comput Ind 85(2017):1–10. https://doi.org/10.1016/j.compind.2016.11.004

Scurati GW, Gattullo M, Fiorentino M, Ferrise F, Bordegioni M, Uva AE (2018) Converting maintenance actions into standard symbols for Augmented Reality applications in Industry 4.0. Comput Ind 98(2018):68–79. https://doi.org/10.1016/j.compind.2018.02.001

Fox S (2010) The importance of information and communication design for manual skills instruction with augmented reality. J Manuf Technol Manag 21(2):188–205. https://doi.org/10.1108/1741039110114369

Zubizarreta J, Aguina I, Amundarain A (2019) A framework for augmented reality guidance in industry. Int J Adv Manuf Technol 102(9):4095–4108. https://doi.org/10.1007/s00170-019-35527-2

Wang Y, Zhang S, Wan B, He W, Bai X (2018) Point cloud and visual feature-based tracking method for an augmented reality-based mechanical assembly system. Int J Adv Manuf Technol 99(9):2341–2352. https://doi.org/10.1007/s00170-018-2575-8

Lu Y, Gao K, Zhang T, Xu T (2018) A novel image registration approach via combining local features and geometric invariants. PLoS One 13(1):e0190383–e0190383. https://doi.org/10.1371/journal.pone.0190383

Schmalstieg D, Hollerer T (2016) Augmented Reality: principles and practice. Addison-Wesley Professional

Yazdi M, Bouwmans T (2018) New trends on moving object detection in video images captured by a moving camera: a survey. Comp Sci Rev 28(2018):157–177. https://doi.org/10.1016/j.cosrev.2018.03.001

Meshgi K, Ishii S (2015) The state-of-the-art in handling occlusions for visual object tracking. IEICE Trans Info Syst E98.D(7):1260–1274. https://doi.org/10.1587/transinf.2014EDR0028

Bhaskar H, Dwivedi D, Dogra DP, Al-Muallal M, Mihaylova L (2015) Autonomous detection and tracking under illumination changes, occlusions and moving camera. Signal Process 115(2017):343–354. https://doi.org/10.1016/j.sigpro.2015.06.003

Dey S, Reilly V, Saleemi I, Shah M (2012) Detection of independently moving objects in non-planar scenes via multi-frame monocular epipolar constraint. In: Fitzgibbon A., Lazebnik S., Perona P., Sato Y., Schmid C. (eds) Computer Vision – ECCV 2012. ECCV 2012. Lect Notes Comp Sci 7576. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-33715-4_62

Zhang Z (1998) Determining the epipolar geometry and its uncertainty: a review. Int J Comput Vision 27(2):161–195. https://doi.org/10.1023/A:1007765809751

Szeliski R (2022) Computer vision: algorithms and applications. Springer Cham. https://doi.org/10.1007/978-3-030-34372-9

Hartley R, Zisserman A (2004) Multiple View Geometry in Computer Vision (2nd ed.). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511811685

Ong SK, Yew AW, Thangaiavel NK, Nee AYC (2020) Augmented reality-assisted robot programming system for industrial applications. Robot Comput Integr Manuf 61(2020)

Xu J, Chen R, Chen H, Zhang S, Chen K (2017) Fast registration methodology for faster assembly of large-scale structure. IEEE Trans Industr Electron 64(1):717–726. https://doi.org/10.1109/TIE.2016.2599140

Xiao H, Duan Y, Zhang Z (2018) Mobile 3D assembly process information construction and transfer to the assembly station of complex products. Int J Comput Integr Manuf 31(1):11–26. https://doi.org/10.1080/0951922X.2017.1356470

Li X, Yi W, Chi H, Wang X, Chan APC (2018) A critical review of virtual and augmented reality (VR/AR) applications in construction safety. Autom Constr 86(2018):150–162. https://doi.org/10.1016/j.autcon.2017.11.003

Pan Z, Liu S, Fu W (2017) A review of visual moving target tracking. Multimedia Tools and Applications 76(16):16989–17018. https://doi.org/10.1007/s11042-016-3647-0

Gao QH, Wan TR, Tang W, Chen L (2019) Object registration in semi-cluttered and partial-occluded scenes for augmented reality. Multimed Tools 78(11):15079–15099. https://doi.org/10.1007/s11042-018-6905-5

Wang Y, Zhang S, Yang S, He W, Bai X, Zeng Y (2017) A LINE-MOD-based markerless tracking approach for AR applications. Int J Adv Manuf Technol 89(5–8):1699–1707. https://doi.org/10.1007/s00170-016-9180-5

Huang P, Cai J, Meng Z, Hu Z, Wang D (2014) Novel method of monocular real-time feature point tracking for tethered space robots. J Aerosp Eng 27(6):04014039. https://doi.org/10.1061/(ASCE)AS.1943-5525.0000367

Wang K, Liu D, Duan G, Hu L, Tan J (2020) A fast object registration method for augmented reality assembly with simultaneous determination of multiple 2D–3D correspondences. Robot Comput Integr Manuf 63(2020)

Goshatsbsy AA (2012) Image registration: principles, tools and methods. Springer, London. https://doi.org/10.1007/978-1-4471-2458-0

Marchand E, Uchiyama H, Spindler F (2016) Pose estimation for augmented reality: a hands-on survey. IEEE Trans Visual Comput Graphics 22(12):2633–2651. https://doi.org/10.1109/tvcg.2015.2513408

Maron H, Dym N, Kezurer I, Kovalsky S, Lipman Y (2016) Point registration via efficient convex relaxation. ACM Trans Graph 35:1–12. https://doi.org/10.1145/2897824.2925913

Xiong Z, Zhang Y (2010) A critical review of image registration methods. Int J Image Data Fusion 1(2):137–158. https://doi.org/10.1007/s11149-010-03027-0

Zhu H, Gao B, Zou K, Li Y, Yuen K, Mihaylova L, Leung H (2019) A review of point set registration: from pairwise registration to groupwise registration. Sensors 19(5):1191. https://doi.org/10.3390/s19051191

Zhang Z, Deriche R, Faugeras O, Luong Q (1995) A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry. Artif Intell 78(1):87–119. https://doi.org/10.1016/0004-3702(95)00022-4

Zamalieva D, Yilmaz A, DAVIS JW (2014) A multi-transformation model for background subtraction with moving cameras. In: Fleet D., Pajdla T., Schiele B., Tuytelaars T. (eds) Computer Vision – ECCV 2014. ECCV 2014. Lect Notes Comp Sci 8689. Springer, Cham. https://doi.org/10.1007/978-3-319-10590-1_52

Ramachandran M, Veeraraghavan A, Chellappa R (2009) CHAPTER 5 - video stabilization and mosaicing, in The Essential Guide to Video Processing:109–140. https://doi.org/10.1007/978-0-387-75656-2.00006-2

Wang J, Gu G, Wan M, Ren K (2018) Dual-correlation transformation for image stitching. J Electron Imaging 27(3):033040. https://doi.org/10.1117/1.JEI.27.3.033040

Fathy ME, Hussein AS, Tolba MF (2011) Fundamental matrix estimation: a study of error criteria. Pattern Recogn Lett 32(2):383–391. https://doi.org/10.1016/j.patrec.2010.09.019
45. Basta T, Emirates U (2013) Does the fundamental matrix define a one-to-one relation between the corresponding image points of a scene? Int J Image Graph 1(3):125–128. https://doi.org/10.12720/JOIG.1.3.125-128
46. Basta T (2014) Is the fundamental matrix really independent of the scene structure? Int J Signal Process, Image Processing and Pattern Recognition 7(5):149–167. https://doi.org/10.14257/ijisp.2014.7.5.13
47. Stojanovic A, Unger M (2007) Robust detection of point correspondences in stereo images. Acta Polytechnica 47(4–5):23–28. https://doi.org/10.14311/956
48. Ben-Artzi G, Werman M, Peleg S (2017) Camera calibration from dynamic silhouettes using motion barcodes. ArXiv 2017(Jan):1–9. https://doi.org/10.48550/arXiv.1506.07866
49. Hassaballah M, Abdelmageid AA, Alshazly HA (2016) Image features detection, description and matching, In: Awad A, Hassaballah M (eds) Image Feature Detectors and Descriptors. Stud Comput Intell 630. Springer, Cham. https://doi.org/10.1007/978-3-319-28854-3_2
50. Stojanovic A, Unger M (2013) A new evaluation criterion for point correspondences in stereo images. In: Adami N, Cavallaro A, Leonardi R, Migliorati P (eds) Analysis, Retrieval and Delivery of Multimedia Content. Lect Notes Electr Eng 158. Springer, New York, NY. https://doi.org/10.1007/978-1-4614-3831-1_11

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.