Contour Primitives of Interest Extraction Method for Microscopic Images and Its Application on Pose Measurement

Fangbo Qin, Fei Shen, Dapeng Zhang, Xilong Liu, and De Xu, Senior Member, IEEE

Abstract—This paper proposes a suite of methods to realize high precision pose measurement in 3-D Cartesian space based on a multicamera microscopic vision system. Since it is inefficient to develop a specific image algorithm for each kind of object and the imaging condition might be unsatisfactory, we propose a method of contour primitives of interest extraction, which allows flexible reconfiguration for novel object image and owns robustness under different imaging conditions. The object is detected in grayscale image based on a template of contour primitives. Edges are extracted according to derivatives along the normal vectors of these contour primitives. The positions and directional derivatives of these edges are used for feature extraction and autofocus, respectively. The point features and line features extracted from multiview images are to measure 3-D vectors and orientations, respectively, based on image Jacobian matrices. Cameras’ linear motions are considered in the imaging model, so that the measurement range is expanded beyond the limitation of microscopes’ shallow depths of field. The affine epipolar constraint and focused planes intersection constraint between cameras are applied to improve the real time performances of image feature extraction and multicamera autofocus, respectively. A series of experiments are conducted to verify the effectiveness of the proposed methods. The root mean square errors of pose measurement are evaluated as 3 μm in position and 0.05° in orientation, while the measurement range is about 5000 μm in position and 20° in orientation.

Index Terms—Geometric constraint, image feature extraction, microscopic vision, pose measurement, precision assembly

I. INTRODUCTION

MICROSCOPIC vision systems have been widely used for noncontact, real-time, and high precision measurement of objects with millimeters or micrometers sizes, such as biological cells, microelectromechanical systems (MEMSs), micro optical devices, microstructures, etc. [1], [2]. A microscopic vision system mainly consists of cameras, microscopes, light sources, and adjusting stages. Microscope offers the advantages of high magnification and ignorable distortion. Telecentric microscope, whose magnification remains constant with object depth changing in a specific range, is preferred in high accuracy measurement for 3-D objects [3]. With the advances of micro-scale technology, the objects are with more precise and complex structures. It remains challenging to improve the accuracy, flexibility, and robustness of microscopic vision measurement. This paper concerns the measurement of pose information, which can be applied in precision assembly, structure inspection, and motion monitoring.

Image feature extraction is an essential issue in microscopic vision measurement, which means transforming the source image into a set of informative features of interest. The feasibility, robustness, and accuracy of measurement highly depend on image feature extraction [26], [27]. Low-level features like edge and corner are detected according to local pixels. Edge feature is popular in pose measurement. Canny algorithm [4] detects single-pixel edges with nonmaximum suppression, and eliminates noise edges by edge tracking with hysteresis. A challenging problem is to extract the object-related features that lie in an image at unknown poses. With various object images to process, extraction methods allowing flexible reconfiguration for novel objects are preferred than those designed for specific tasks. Hough transform-based methods are widely used for detection of line, circle, and even arbitrary shape [5], [6]. Shark et al. [7] proposed the feature matching method based on line segments. Both the Hough transform-based and line segments-based methods rely on the preprocessing step of edge extraction. However, microscopic image is prone to defocus, so that the edge extraction is not robust as a preliminary step. Grayscale template-based matching methods search for object in grayscale image [8]. The template contains a region of pixels and can describe complex objects. However, grayscale template matching is computationally expensive. The fast affine template matching (fast-match) algorithm accelerated the matching by the random sampling and the branch-and-bound scheme [9]. The shape-based matching (SBM) method was presented in [10]. It constructs a point set as the shape model by edge extraction from template image, and searches for the object based on the consistency of image gradient. SBM is robust against illumination change and occlusion. However, the shape model in SBM might involve edges that are produced by features of no
interest, such as texture. Although changes of magnification and viewpoint are minor in microscopic vision, bad imaging conditions, such as overlap, illumination change, occlusion, blur, and weak feature, often cause failures of image feature extraction. In addition, a microscopic image is with large size, and algorithms need be fast for real time operation.

Autofocusing is necessary to a microscopic camera, which realizes acquiring clear images by adjusting its object distance automatically. It is implemented by translating the camera along a linear motion stage to search for the position where the focus measure (FM) is maximum. The FM versus camera position curve is expected to have a sharp global peak at the exact best focus position and no false local peaks at other positions. The mountain climbing algorithm is a broadly adopted searching strategy with high efficiency [11]. Sun et al. [12] presented a systematic investigation of 18 FM algorithms, among which the normalized variance algorithm provided the best overall performance. However, autofocus might be not accurate for a 3-D object, because its visible parts locate at different depths. Only the part of interest need be focused on, which depends on the selection of image region.

As is different from conventional lens, microscope is with the limitations of narrow field of view and shallow depth of field, which are challenges to microscopic vision. Monocular microscopic camera is suitable for planar measurement [13]–[15]. Wang et al. [16] presented an automated 3-D micro-grasping method, in which the microscopic camera offered 2-D position feedback. 3-D information of object could be recovered from multiview images using stereo vision [17], [18]. Yamamoto and Sano [19] used a stereoscopic microscope to measure the needle tip’s 3-D position in a micromanipulation system. In [20], a multicamera visual tracking system was developed for microassembly, which provided six degree-of-freedom (DOF) pose feedback of MEMS components in real time. It relied on the CAD model and initial value estimation. Its position measurement accuracy decreased with the object’s motion distance. Shen et al. [21] and Liu et al. [22] implemented pose alignment control based on visual feedbacks of multiple microscopic cameras in the 3-D precision assembly tasks. The objects’ relative poses were measured using image Jacobian matrices and image features. However, their relative position measurements were limited within the optical depth of field, i.e., submillimeter range. In many cases, the relative depth between two objects exceeds the optical depth of field, so that the camera needs to be translated to focus on them sequentially. The measurement accuracy cannot be guaranteed when the camera motion is not considered. In addition, the relative attitudes between objects were obtained in the joint space. However, some tasks require attitudes that are uniformly represented in 3-D Cartesian space.

The motivation of this paper is to design a microscopic vision system for high precision pose measurement in 3-D Cartesian space. A method of contour primitives of interest extraction (CPIE) is proposed to obtain object-related features from grayscale image in real time. The object detection and edge extraction are based on the object’s contour primitives of interest, which do not involve parts that are irrelevant to measurement. The method can output image features and contour sharpness for pose measurement and autofocus, respectively. Given a novel object, user can reconfigure the method using a drawing interface, instead of reprogramming for it. The 3-D vector and orientation measurement methods are proposed, which are based on the image Jacobian matrices of multiple telecentric cameras. Camera motions along linear stages are considered to expand the measurement range beyond the limitation of depth of field. The affine epipolar constraint and focused planes intersection constraint between cameras are applied to reduce the time cost of corresponding feature extraction and multicamera autofocus, respectively. The effectiveness of the proposed methods is verified by the experiments on a precision assembly system. The main contributions of this paper are as follows.

1) A robust feature extraction method allowing reconfiguration for novel object images is proposed.
2) The pose measurement is realized in a range larger than the depth of field.
3) The geometric constraints between cameras are utilized to improve the measurement efficiency.

The remainder of this paper is organized as follows. Section II is the system overview. The CPIE method is proposed in Section III. Section IV describes the imaging model and the pose measurement methods. The geometric constraints between cameras and their applications are given in Section V. Section VI provides the experimental results. Finally, this paper is concluded in Section VII.

II. System Overview

The precision assembly system consists of three telecentric microscopic cameras, three linear motion stages, and two manipulators, as shown in Fig. 1(a). Each camera is mounted on a linear motion stage, whose translation axis is approximately parallel to the camera’s optical axis. The linear stage is used to adjust the camera’s object’s distance, so that features at different depths can be clearly imaged sequentially. A ring light and a backlight are installed for each camera, so that either surface image or silhouette image can be acquired. The manipulator 1 is with three translational DOFs. The manipulator 2 has three rotational DOFs and one vertical translational DOF.

The camera coordinates \( \{C_1\}, \{C_2\}, \) and \( \{C_3\} \) are established at the top left corners of the charge-coupled devices of the cameras 1, 2, and 3, respectively. Their \( z_c \) axes are parallel to the cameras’ optical axes and point to the scene. Their \( x_c \) axes are corresponding to the horizontal axes of the image coordinates. The world coordinates \( \{W\} \) are established to be identical with the manipulator 1 coordinates, whose origin is at the manipulator 1’s base. The manipulator 2 coordinates are established at the manipulator 2’s base, whose axes are parallel to those of \( \{W\} \).

The block diagram of the microscopic vision system is given in Fig. 1(b). It consists of autofocusing, image capture, feature extraction, and pose measurement modules. The autofocusing module is implemented via mountain-climbing search according to the contour sharpness given by the feature
extraction module, or via moving the camera directly to the
position given by the focused planes intersection constraint
when another camera finishes focusing on the same object.
After autofocus, the camera captures the object image. The
feature extraction module is used to obtain image features and
contour sharpness from source image with the CPIE method,
whose template is configured by user in advance. The affine
epipolar constraint provides restricted search range of corre-
sponding feature to reduce the execution time of CPIE. The
pose measurement module calculates the vector formed by two
scene points using the corresponding point features in differ-
ent views. It can also calculate the line’s orientation with the
corresponding line features in different views.

III. CONTOUR PRIMITIVES OF
INTEREST EXTRACTION

A common observation about many object images is that
their shapes can be represented by contour primitives, such
as lines segments and circular arcs. The object detection
can be based on the contour primitives of interest, without involv-
ing other features like texture and irregular parts. Furthermore,
image features and contour sharpness can be obtained using the extracted pixels of these contour primitives. As shown in
Fig. 2, the CPIE method includes four steps: 1) template cre-
ation; 2) object detection; 3) edge extraction; and 4) feature
extraction.

A. Template Creation

The template of an object is a set of specific contour prim-
itives, as shown in Fig. 2(a). As an example, the template in
Fig. 2(a) contains four line segments and one circular arc. The
contour primitives are represented in the template coordinates
{TCP\}. Line segment \(L_s\) and circular arc \(A_c\) are two typical
contour primitives. The points on them are given by

\[ L_s : p(l) = \begin{bmatrix} u_l + d_{ul}l \\ v_l + d_{vl}l \end{bmatrix}, \quad l \in [0, l_0] \] (1)

\[ A_c : p(\varphi) = \begin{bmatrix} u_a + r \cos \varphi \\ v_a + r \sin \varphi \end{bmatrix}, \quad \varphi \in \left[ \begin{bmatrix} \varphi_0, \varphi_1 \end{bmatrix}, \begin{bmatrix} \varphi_1, \varphi_0 \end{bmatrix}, \varphi_0 > \varphi_1 \right] \] (2)

where \([u_l, v_l]^T\), \([d_{ul}, d_{vl}]^T\), and \(l_0\) are the start-point, direction
vector, and length of \(L_s\), respectively. \([u_a, v_a]^T\), \(r\), \(\varphi_0\), and
\(\varphi_1\) are the center, radius, start-angle, and end-angle of \(A_c\),
respectively.

Moreover, each contour primitive indicates the border of
different brightness between its two sides. The positive side
and negative side are defined to describe an outer point’s rel-
ative position to a contour primitive. Given a point \([u, v]^T\), it
is on the positive side of \(L_s\) if \([-d_{ul}, d_{vl}][u - u_l, v - v_l]^T > 0\),
and on the negative side of \(L_s\) if \([-d_{ul}, d_{vl}][u - u_l, v - v_l]^T < 0\).
Similarly, the point is on the positive side of \(A_c\) if \([u - u_a]^2 +
(v - v_a)^2 - r^2 \cdot \text{sgn} (\varphi_1 - \varphi_0) > 0\), and on the negative side
of \(A_c\) if \([u - u_a]^2 + (v - v_a)^2 - r^2 \cdot \text{sgn} (\varphi_1 - \varphi_0) < 0\). The
template creation should satisfy that the positive side of each
contour primitive is brighter than the negative side.
The template $\{L_{e1}, L_{e2}, \ldots, L_{em}, A_{e1}, A_{e2}, \ldots, A_{em}\}$ is created based on the prior knowledge about the object contour. User’s manual selection on a drawing interface is a direct and convenient approach for template creation. Given an object image, the user specifies the line segments, circular arcs, and origin point of $\{T_{CP}\}$ on this image. The orientation of $\{T_{CP}\}$ is identical with that of the image coordinates. The user-specified $L_{e}$ and $A_{e}$ are recorded as the parameter lists $\{u_{i}, v_{i}, d_{u}, d_{v}, l_{i}\}$ and $\{u_{i}, v_{i}, r, \psi_{i}, \varphi_{i}\}$, respectively. The created template can be used to detect the objects of the same type.

B. Object Detection

When the template lies in the image coordinates after the transformation $T$, the pixels near the transformed contour primitives of interest are analyzed to detect the object.

Given a point $p$ on the transformed template, its position is represented in the template coordinates. The points $p_{+}$ and $p_{-}$ are on the positive and negative sides of $p$, respectively, and their distances to $p$ both equal $\rho$. The vector pointing from $p_{+}$ to $p_{-}$ is along the contour primitive’s normal direction at $p$. The positions of $p_{+}$ and $p_{-}$, represented in the template coordinates, are given by

$$
\begin{align*}
L_{e} : & \quad \{p_{+}(l) = p(l) + [−\rho d_{u}, \rho d_{v}]^{T} \} \\
& \quad \{p_{−}(l) = p(l) − [−\rho d_{u}, \rho d_{v}]^{T} \}
\end{align*}
$$

$$
\begin{align*}
A_{e} : & \quad \{p_{+}(\varphi) = p(\varphi) + [\text{sgn}(\psi_{i} - \varphi_{i})\rho \cos \varphi, \rho \sin \varphi]^{T} \} \\
& \quad \{p_{−}(\varphi) = p(\varphi) − [\text{sgn}(\psi_{i} - \varphi_{i})\rho \cos \varphi, \rho \sin \varphi]^{T} \}
\end{align*}
$$

As the transformation from the template coordinates to the image coordinates is $T$, these points’ positions are represented as $p' = T(p)$, $p_{+}' = T(p_{+})$, and $p_{−}' = T(p_{−})$ in the image coordinates, respectively. With the telecentric camera, the scale change of object image is minor. Therefore, the transformation $T$ is only determined by the two translation parameters $u_{t}$, $v_{t}$ and the rotation angle $\theta$. Namely

$$
\begin{align*}
p' = T(p) = \begin{bmatrix}
\cos \theta & −\sin \theta \\
\sin \theta & \cos \theta
\end{bmatrix} p + \begin{bmatrix}
u_{t} \\
v_{t}
\end{bmatrix}.
\end{align*}
$$

$N_{s}$ points $p_{i}'$ ($i = 1, 2, \ldots, N_{s}$) are sampled on the transformed template with an equal length interval $s$, as shown in Fig. 2(b). The matching ratio under the transformation $T$ is given by

$$
H_{T} = \frac{1}{N_{s}} \sum_{i=1}^{N_{s}} h(I(p_{+}'), I(p_{−}'), \tau_{c})
$$

$$
h(\tau) = \begin{cases} 
1, & \tau > 0 \\
0, & \tau \leq 0
\end{cases}
$$

where $I(p)$ is the intensity at $p$. $\tau_{c}$ is a positive contrast threshold.

If the transformed template matches with the object, the sampling points $p_{i}'$ will satisfy $I(p_{+}') - I(p_{−}') > \tau_{c}$. Considering the occlusion, defect and overlap on the object contour, the matching ratio is allowed to be smaller than one, but is expected to be larger than the threshold $\eta \in [0.5, 1]$, when the transformed template matches with the object.

In addition, a higher intensity contrast between the positive and negative sides of the transformed template gives more confidence in object detection, which is evaluated by the contrast-weighted score

$$
F_{T} = \frac{1}{N_{s}} \sum_{i=1}^{N_{s}} (l(p_{+}') - l(p_{−}')) h(I(p_{+}') - I(p_{−}'), \tau_{c}).
$$

The result of object detection is the best transformation $T$ that maximizes $F_{T}$ and satisfies $H_{T} > \eta$. As an example, Fig. 2(b) gives the object detection result. The green dots denote the sampling points $p'_{i}$. The points $p_{+}'$ and $p_{−}'$ associated with $p'_{i}$ are marked by red and yellow dots, respectively.

The searching for the best transformation is conducted in the transformation parameters space. The searching step length is $\rho$ in the $u_{t} - v_{t}$ space and $\omega$ in the $\theta$ space. Restriction of search ranges can reduce the time cost on object detection. Moreover, the matching ratio under each transformation is obtained by visiting the $N_{s}$ sampling points in a random order and count the sampling points satisfying $I(p_{+}') - I(p_{−}') > \tau_{c}$. The visiting of the sampling points can be stopped as soon as more than $(1 - \eta)N_{s}$ visited sampling points satisfying $I(p_{+}') - I(p_{−}') \leq \tau_{c}$ have been found, because it is certain that the matching ratio under this transformation is lower than the minimum threshold. Thus, the computation time of object detection is reduced.

In addition, it should be noticed that a scale factor of template is considered since the object’s image size changes with its depth if CPIE is used in traditional vision systems.

C. Edge Extraction

Edges of the contour primitives of interest are extracted based on the object detection result. Instead of searching for the edges globally, we take advantage of the fact that the intensity changes sharply at the edge position along the contour’s normal direction. Normal derivative is the directional derivative along the normal direction vector. Hence, the edges can be extracted by searching for the points giving local maxima of normal derivative within the neighborhood of each contour primitive.

Given a point $p$ on the transformed template, the unit normal vector of the contour primitive at this point is $v_{n}$, as shown in Fig. 3. The points $p_{+}$ and $p_{−}$ associated with $p$ form the line segment $P_{+}P_{−}$. The edge $p_{e}$ on $P_{+}P_{−}$ is the point giving the maximum directional derivative $d_{m}$ along $v_{n}$. Namely

$$
p_{e} = \arg \max_{p \in P_{+}P_{−}} \nabla_{v_{n}} I(p)
$$

$$
d_{m} = \nabla_{v_{n}} I(p_{e}).
$$

In addition, only the point satisfying $d_{m} > \tau_{d}$ is identified as an edge, where $\tau_{d}$ is a positive directional derivative threshold. Thus, a list of edges is obtained for each contour primitive of the transformed template.
The contour sharpness, determined by focus, is decisive to the measurement accuracy. If the contour primitives of interest are out of focus, the detected edges might be inexact. Therefore, the search-based autofocus is implemented to maximizing the sharpness of contour primitives of interest, instead of maximizing the traditional FM of an image region. The average of the edges’ normal derivatives are used to evaluate the contour sharpness

\[ S_c = \frac{1}{N_e} \sum_{i=1}^{N_e} d_{mi} \]  

(11)

where \( N_e \) is the number of extracted edges. \( d_{mi} \) is the normal derivative of the \( i \)th edge. Compared to other FM algorithms, the evaluation of contour sharpness is more concentrated on the interesting contour, without involving other irrelevant features.

After edge extraction, the contour primitives of interest are represented by several lists of edges, and the contour sharpness indicates the focus status. An example is given in Fig. 2(c). When the contour sharpness is large enough, the edges are exact for the next feature extraction step.

D. Feature Extraction

The image features like points and lines can be determined using the geometric relationship of the contour primitives of interest. First, the lists of edges are used for shape fitting to acquire the exact equations of these contour primitives. There might exist defect, occlusion, and overlap on the object contour, which produce false edges. Therefore, the robust shape-fitting algorithm with \( M \)-estimator [23] is adopted, in which the influence of outlying edges is suppressed. Finally, the fitted equations are used to calculate the features of interest. As an example, Fig. 2(d) shows the extracted features. The obtained point and line feature are marked by magenta.

IV. IMAGING MODEL WITH LINEAR MOTION AND POSE MEASUREMENT METHOD

The point and line features obtained from multiview microscopic images are used to calculate the vectors formed by points and the orientations of lines in 3-D Cartesian space. To overcome the problem that the measurement range is limited by the shallow depth of field, we introduce the camera’s linear motion into the imaging model and the pose measurement method. Thus, the image features with different object depths can be used for pose measurement without loss of accuracy.

A. Imaging Model With Linear Motion

Telecentric microscopes are adopted due to their advantages of good resolution, ignorable distortion, and nonperspective effect. The object-side telecentric lens owns an aperture stop exactly at the lens focal point. With this configuration, only ray cone whose principal axis is parallel to the lens optical axis is collected by the lens pupil. Thus, the telecentric lens produces orthographic projection, so that the magnification is independent of object depth. Ignoring the lens distortion, the imaging model is

\[ p = \begin{bmatrix} \alpha & \gamma & 0 \\ 0 & \beta & 0 \\ 1 & 0 & 1 \end{bmatrix} C P + \begin{bmatrix} u_0 \\ v_0 \end{bmatrix} \]  

(12)

where \( C P = [C_x, C_y, C_z]^T \) is a scene point in the camera coordinates \( C \), and \( p = [u, v]^T \) is its image point. \( \alpha, \beta, \gamma, u_0, \) and \( v_0 \) are the intrinsic parameters. The imaging model can be considered as a combination of an orthographic projection onto the \( x_c - y_c \) plane of \( C \) and an affine transformation into the image coordinates, as shown in Fig. 4.

Focused plane is a virtual plane that is perpendicular to the optical axis and lies at the best focus depth \( D_f \). A scene point on the focused plane produces an image point. A scene point out of the focused plane is imaged as a blur circle, whose radius increases with the out-of-plane distance. Depth of field is the depth range that provides acceptable blur circle. Although (12) is independent of object depth, only scene points locating within the depth of field can be clearly imaged. Therefore, the camera’s translation is required for its object distance adjustment. \( v_c \) denotes the positive direction of the linear motion stage’s axis, as shown in Fig. 4. As the camera is translated by the linear motion stage, the focused plane is translated by the same displacement.

The extrinsic parameters are denoted by the rotation matrix \( CW \) and the translation vector \( ct \). The latter is dependent on the camera’s position \( p_c \) on the linear motion stage, as expressed as

\[ ct = \begin{bmatrix} t_1 + p_1 c_1 \\ t_2 + p_2 c_2 \\ t_3 + p_3 c_3 \end{bmatrix} \]  

(13)

where \( t_1, t_2, \) and \( t_3 \) are constants. \([c_1, c_2, c_3]^T\) is a normalized direction vector depending on \( v_c \). Considering the extrinsic parameters, the imaging model with linear motion is given by

\[ P = \begin{bmatrix} \alpha & \gamma & 0 \\ 0 & \beta & 0 \\ 1 & 0 & 1 \end{bmatrix} WP + \begin{bmatrix} u_0 \\ v_0 \end{bmatrix} \]  

\[ WP = \begin{bmatrix} W_x, W_y, W_z \end{bmatrix} \]  

\[ \begin{bmatrix} r_{11}, r_{12}, r_{13} \\ r_{21}, r_{22}, r_{23} \end{bmatrix} \]  

\[ \begin{bmatrix} u_0 \\ v_0 \end{bmatrix} \]  

\[ D_1 \]  

(14)

where \( WP = [W_x, W_y, W_z]^T \) is the scene point represented in the world coordinates \( [W] \). \([r_{11}, r_{12}, r_{13}] \) and \([r_{21}, r_{22}, r_{23}] \) are the first two rows of \( CW \). If \( v_c \) is not strictly parallel to the \( z_c \) axis of \( C \), \( t_1 \) and \( t_2 \) are nonzero, and the camera’s translation on the linear stage will cause the position offset of image points.
B. Three-Dimensional Vector Measurement

The 3-D vectors can describe both the relative positions and absolute attitudes. They are formed by scene points like centers, intersection points, corners, tips, markers, etc. A 3-D vector can be measured if it is formed by two scene points that are commonly imaged by \( n \geq 2 \) cameras.

Given two scene points \( P_1 \) and \( P_2 \), a 3-D vector \( \Delta P = P_2 - P_1 \) is formed in \( [W] \). If the relative depth of \( P_1 \) and \( P_2 \) is smaller than the camera’s depth of field, their image points can be obtained by the camera at one time. Otherwise, the camera should be translated to acquire the clear images of \( P_1 \) and \( P_2 \) sequentially. Without loss of generality, the image points of \( P_1 \) and \( P_2 \), labeled as \( p_1 \) and \( p_2 \), are obtained when the camera locates at the positions \( r_{c1} \) and \( r_{c2} \), respectively. Thus, the image vector \( \Delta p = p_2 - p_1 \) is dependent on both the scene vector \( \Delta P \) and the camera translation \( \Delta r_c = r_{c2} - r_{c1} \).

\[
\Delta p = J \Delta P + J_c \Delta r_c
\]

where \( J \) is the image Jacobian matrix. \( J_c \) is called the image offset matrix, which determines the image offset caused by the camera’s linear motion.

For monocular imaging, the depth information is lost in the orthogonal projection. With \( n \geq 2 \) cameras viewing the same scene vector from different directions, we have

\[
\begin{bmatrix}
\Delta p_1 - J_{1i} \Delta r_{ci} \\
\Delta p_2 - J_{2i} \Delta r_{ci} \\
\vdots \\
\Delta p_n - J_{ni} \Delta r_{ci}
\end{bmatrix} =
\begin{bmatrix}
J_1 \\
J_2 \\
\vdots \\
J_n
\end{bmatrix} \Delta P
\]

where \( J_i, J_{ci}, \Delta p_i, \) and \( \Delta r_{ci} \) are the image Jacobian matrix, image offset matrix, image vector, and translation length of the camera \( i (i = 1, 2, \ldots, n) \). \( J_i \) and \( J_{ci} \) are calibrated beforehand. \( \Delta p_i \) is calculated by the two point features extracted from images. \( \Delta r_{ci} \) is read from the linear motion stage controller. Hence, \( \Delta P \) can be estimated using the least squares method (LSM), as given by

\[
\Delta P = J^\dagger \begin{bmatrix}
J_1 \\
J_2 \\
\vdots \\
J_n
\end{bmatrix} \begin{bmatrix}
\Delta p_1 - J_{1i} \Delta r_{ci} \\
\Delta p_2 - J_{2i} \Delta r_{ci} \\
\vdots \\
\Delta p_n - J_{ni} \Delta r_{ci}
\end{bmatrix}
\]

where \( \dagger \) is the pseudoinverse symbol.

C. Orientation Measurement

The object’s orientation in 3-D Cartesian space can be indicated by lines like center axis and straight edges. The orientation of the line can be measured if the line is imaged by \( n \geq 2 \) cameras. A segment on the line within the field of view and depth of field is clearly viewed by each camera. Generally, the segments imaged by different cameras are independent to each other, but they provide constraints on the same orientation.

Given a line in \( [W] \), its normalized direction vector is denoted by \( v \). The line’s projection in the image is \( l_i \), whose 2-D normal vector is denoted by \( n \). Using the image Jacobian matrix, the image vector projected from the scene vector \( v \) is given by \( Jv \), which is parallel to \( l_i \). Thus, we have

\[
n \cdot Jv = 0.
\]

With \( n \geq 2 \) cameras observing the same line from different directions, the constraints are constructed by

\[
Uv = \begin{bmatrix}
n_1^T J_1 \\
n_2^T J_2 \\
\vdots \\
n_n^T J_n
\end{bmatrix}v = \begin{bmatrix}0 \\
\vdots \\
0\end{bmatrix}
\]

where \( n_i \) is the normal vector of the line’s image in camera \( i (i = 1, 2, \ldots, n) \). The estimation of \( v \) is solved by the optimization

\[
\min_v \|Uv\|^2, \quad \text{s.t. } \|v\| = 1.
\]

The optimal estimation of \( v \) is the eigenvector that is associated with the smallest eigenvalue of \( U^TU \) and the minimum of \( \|Uv\|^2 \) equals this smallest eigenvalue [24].

V. CORRESPONDING FEATURE EXTRACTION AND AUTOFOCUS WITH GEOMETRIC CONSTRAINTS

Since an object should be clearly viewed by at least two cameras in pose measurement, we investigate the geometric constraints between cameras. First, the corresponding image points in two views satisfy the epipolar constraint, which helps for corresponding feature extraction. Second, the two camera’s positions on the linear motion stages satisfy the focused planes intersection constraint to acquire two clear views, which aids in multicamera autofocus.

A. Corresponding Feature Extraction With Affine Epipolar Constraint

As shown in Fig. 5, a point \( p_{1i} \) on the image plane 1 back-projects to a ray in 3-D Cartesian space. This ray is projected as the epipolar line \( l_{12} \) on the image plane 2. \( p_{1i} \)’s corresponding image point \( p_{12} \) lies on the epipolar line \( l_{12} \), which is called the epipolar constraint. Telecentric camera is considered as affine camera, for its projection center lies at infinity. Given a scene point that produces two image points \( p_{1i} = [u_{1i}, v_{1i}] \) and \( p_{12} = [u_{12}, v_{12}] \) in the two telecentric cameras 1 and 2, respectively, the affine epipolar constraint [25] yields

\[
a u_{1i} + b v_{1i} + c u_{12} + d v_{12} + e = 0
\]

where \( a, b, c, d, \) and \( e \) are constants for fixed cameras.
We assume that the parameters in (21) are determined when the two cameras’ positions on their linear motion stages are both zero. When the two cameras are translated to the positions \( p_{11} \) and \( p_{21} \), the corresponding image points are obtained as \( p_{11} \) and \( p_{21} \). Then \( p_{11} = J_{c1}p_{11} \) and \( p_{21} = J_{c2}p_{21} \) are considered as the corresponding image points assuming the two cameras’ positions are both zero, which satisfy
\[
a(u_{11} - j_{11}p_{11}) + b(v_{11} - j_{12}p_{11}) + c(u_{12} - j_{21}p_{22}) + d(v_{12} - j_{22}p_{22}) + e = 0 \quad (22)
\]
where \( J_{c1} = [j_{11}, j_{12}]^T \) and \( J_{c2} = [j_{21}, j_{22}]^T \) are the image offset matrices of the camera 1 and 2, respectively. Therefore, if the image point \( p_{11} \) is obtained, its corresponding point \( p_{21} \) is expected to lie on the epipolar line defined by \( [a, b, c, d, e] = 0 \) and \( [u_{11}, v_{11}] \). Similarly, the epipolar line associated with \( p_{21} \) is \( [a, b, c, d, e] = 0 \) and \( [u_{12}, v_{12}] \).

With the affine epipolar constraint, the position searching range for corresponding feature can be restricted to the epipolar line instead of the whole image, which can improve the real time performance of the feature extraction module.

**B. Autofocus With Focused Planes Intersection Constraint**

The shallowness depth of field of microscope provides the constraint on the object’s depth. If a scene point is focused on, it lies on the focused plane. As shown in Fig. 5, if two cameras focus on the same scene point, the two focused planes’ intersection line passes this scene point, which is called the **focused planes intersection constraint**.

For the two cameras 1 and 2, a scene point’s position is represented as \([C^1_x, C^1_y, C^1_z]^T \) and \([C^2_x, C^2_y, C^2_z]^T \) in the two camera coordinates \( [C_1] \) and \( [C_2] \), respectively. The transformation from \( [C_1] \) to \( [C_2] \) is indicated by the rotation matrix \( C^2R_{C_1} \) and translational vector \( C^2t_{C_1} \). Note that \( C^2t_{C_1} \) is dependent on the two cameras’ positions \( p_{11} \) and \( p_{22} \).

When the camera 1 focuses on the scene point, the image point \([u_{11}, v_{11}] \) is obtained. Then we have
\[
\begin{align*}
\begin{bmatrix}
C^1_x \\
C^1_y \\
C^1_z \\
\end{bmatrix} &= \begin{bmatrix}
a_1 & \gamma_1 & 0 \\
0 & \beta_1 & \gamma_1 \\
\end{bmatrix}^{-1} \begin{bmatrix}
u_{11} \\
u_{01} \\
\end{bmatrix} \\
\end{align*}
\]
\[
C^2z = Df_1 \\
(23)
\]
where \( a_1, \beta_1, \gamma_1, u_{01}, \) and \( v_{01} \) are the intrinsic parameters of the camera 1. \( Df_1 \) is the best focus depth of the camera 1.

If the camera 2 also focuses on the same scene point, the scene point’s depth in \( [C_2] \) is
\[
C^2z = r_{31}'C^1x + r_{32}'C^1y + r_{33}'C^1z + t_3' + t_{11}'p_{11} + t_{22}'p_{22} = Df_2 \\
(24)
\]
where \([r_{31}', r_{32}', r_{33}'] \) is the last row of \( C^2R_{C_1} \) and \( t_3' = t_{11}'p_{11} + t_{22}'p_{22} \) is the last entry of \( C^2t_{C_1} \) and \( t_{11}' \) and \( t_{22}' \) are determined by the two linear motion stages’ axis directions in \([C_2] \).

Since the linear motion stage 2’s axis is approximately parallel to the camera 2’s optical axis, we have \( t_{22}' \neq 0 \). Combining (23) and (24), the camera 2’s best focus position, which satisfies the focused planes intersection constraint, is given by
\[
p_{22}' = a'u_{11} + b'v_{11} + c'p_{11} + d' \\
(25)
\]
where \( a', b', c', \) and \( d' \) are constants. \( p_{22}' \) is predicted best focused position of camera 2.

It is concluded that the second camera’s best focus position for a scene point can be determined by the image point and the best focus position of the first camera. Therefore, if one camera finishes autofocus or remains in-focus, the other camera can be directly translated to the best focus position given by (25), which costs much less time than the search-based autofocusing methods. However, if neither of the cameras has the object in focus, at least one camera needs to implement search-based autofocusing.

**VI. EXPERIMENTS AND RESULTS**

**A. Hardware Configuration**

The precision assembly system was built based on the design in Section II, as shown in Fig. 6. The three Basler piA2400-17gm cameras (image size: 2448 \( \times \) 2050 pixel; pixel size: 3.45 \( \mu \)m \( \times \) 3.45 \( \mu \)m; and 8-bit grayscale) were mounted with the Myutron MGTL10VC telecentric microscopes (magnification: \( x \)1.0; optical resolution: 2.5 \( \mu \)m; TV distortion: <0.01%; and depth of field: 430 \( \mu \)m). The three Sugura KS102-100 linear motion stages provided the translations with the 1 \( \mu \)m resolution and \( \pm 0.3 \mu \)m repeatability. The manipulator 1 was constructed by the Kohzu YA10A-L1 and ZA10A-X1T stages, whose resolution and repeatability were 1 \( \mu \)m and \( \pm 0.5 \mu \)m, respectively. The rotational axes of the manipulator 2 were constructed by Kohzu SA07A-RL and RA07A-W, whose resolution and repeatability were 0.004\(^\circ\) and \( \pm 0.003\(^\circ\)\), respectively. The image processing algorithms was run on the computer with an Intel Core i5-3337U 1.80 GHz CPU.

**B. Contour Primitives of Interest Extraction**

1) **CPIE on Images of Cylindrical and Conical Objects:**

The cylindrical and conical micro objects were shown in the top-right corner of Fig. 6. The upper object 1 was cylindrical, whose radius and height were about 650 \( \mu \)m and 2500 \( \mu \)m, respectively. The lower object 2 had a conical body and a semicircle groove at its top end. The objects 1 and 2 were gripped by the vacuum grippers of the manipulators 1 and 2, respectively. The scene point that indicated the object 1’s position was the end face’s center. The images of the object 1 in the cameras 1 and 2 were labeled as \( I_1 \) and \( I_2 \), respectively. The line that indicated the object 2’s orientation was the
conical body’s center axis. The images of the object 2 in the cameras 1 and 2 were labeled as $I_3$ and $I_4$, respectively.

First, we created the templates by manual selection on the drawing interface. For $I_1$, five line segments were specified to create the template 1 $\{L_{a1}[−701, −184, 0.999984, −0.005682, 176], L_{a2}[−219, −185, 0.999986, 0.005236, 191], L_{a3}[−3, −165, 0.051215, 0.998688, 39], L_{a4}[−2, 126, 0.0, 1.0, 38], L_{a5}[−26, 189, −0.999999, −0.001488, 672]\}$. For $I_2$, a circular arc was specified to create the template 2 $\{A_c[0, 0, 185, 0°, 360°]\}$. $I_3$ and $I_4$ shared the template 3 $\{L_{a1}[−354, 158, 0.241797, −0.970327, 318], L_{a2}[275, −162, 0.254643, 0.967035, 330]\}$.

The parameters of CPIE were set and selected as follows. The following three parameters used their default values: 1) angle step $\omega = 1°$; 2) contrast threshold $\tau_c = 20$; and 3) normal derivative threshold $\tau_d = 10$. The position step $\rho$ and sample length interval $s$ were selected in the trial and error manner to achieve satisfying performance. $\rho$ determined the position resolution of object detection, and was selected according to the object size. $s$ determined the sampling point number on the object contour. A smaller $s$ gave higher accuracy but larger time cost. Their default values were $\rho = 8$ and $s = 16$. Matching ratio threshold was set as the maximum possible ratio of occlusion, defect, or overlap on object contour, whose default value was $\eta = 0.9$. The position search range was the whole image with the size of $2448 \times 2050$ in pixel. The angle search range was restricted to $[−10°, 10°]$. Because the template 2’s shape was isotropic, the angle search range for it was specially restricted to $[0°,]$

The results of CPIE on $I_1$–$I_4$ were given in Fig. 7. The execution time of CPIE on $I_1$, $I_2$, $I_3$, and $I_4$ was 46.8 ms, 15.4 ms, 36.7 ms, and 37.4 ms, respectively. The CPIE method’s real-time performance satisfied the demand of precision assembly task. Note that for the four images only the templates and the angle search range were reconfigured in the CPIE method.

The contour sharpness given by the edge extraction step of CPIE was used as the FM for autofocusing. We translated the each camera for 20 times with the 100 $\mu$m step. At each position, the contour sharpness of $I_1$ and $I_2$ were recorded.

For the comparison, the normalized variance [12] of the object region was also calculated at each position. The recorded FMs were normalized to the range $[0, 1]$. The normalized FM versus camera position curves were shown in Fig. 8.

In Fig. 8(a), the FM curve using contour sharpness, given by CPIE, is with the steeper slope and the sharper peak compared to the one using the normalized variance algorithm, which is better for peak searching. In Fig. 8(b), the FM curve using normalized variance exhibits a local peak. The local peak was caused by that the end faces of the gripper and object 1 located in the same region but at different depths. In comparison, the CPIE method provided accurate FM for the 3-D object’s outline.

What is more, the CPIE method showed the robustness against image defocus, which worked even when the object depth changed with a range larger than the depth of field. Note that the CPIE method failed to find the object 1 in $I_2$ when the camera 2 position deviated more than 500 $\mu$m from the best focus position, because the blurred bright texture mixed with the blurred object contour.

2) Comparison Experiments of Object Detection: Another two object detection methods were selected for the comparison: the SBM [10] and fast affine template matching (fast-match) [9] on these images. Based on the object poses given by these two methods, the edge and feature extraction steps, which were the same with those in the CPIE method, were employed to obtain the precise image features. The performance of the methods including CPIE, SBM, and fast-match was examined on different objects under five imaging conditions. The five types of images are shown in Fig. 9, and their imaging conditions are described as follows.

Type 1: The object was a micro ball whose material was semitransparent. After it was assembled in the cavity, its image overlapped with the thin film’s image.

Type 2: The object was a micro ball. After it was adsorbed by the manipulator’s vacuum gripper, an upper part of the ball was occluded by the gripper.

Type 3: The object was a square pad with a reflective surface. Its image’s brightness changed significantly when the object’s attitude changed with a small angle under the ring light.

Type 4: The object was a polygonal end face. Its image exhibited weak contour and bright surface textures.
Type 5: The object was a cylindrical cavity. Its image contrast was lower than the template image’s contrast. The reflected light on the surface was also different.

The parameters of the CPIE, SBM, and fast-match methods were configured as follows.

1) CPIE Method: The default parameters were the same with those in Section VI-B. For the images of the types 1 and 2, we selected $\rho = 4$ and $s = 4$, due to the micro balls’ small sizes. The minimum threshold of matching ratio was set as $\eta = 0.66$, while the value range of matching ratio was $[0, 1]$. The templates for CPIE were created by manual drawing on the template images.

2) SBM Method: The shape models were created by applying the Canny algorithm on template images. The position and angle steps were the same with those in CPIE. The minimum threshold of similarity measure was set as 0.33, while the value range of similarity measure was $[-1, 1]$.

3) Fast-Match Method: The photometric invariance is considered in the matching distances evaluation. Only 2-D rigid transformations were configured as candidate transformations. The two parameters determining the matching accuracy were $\delta = 0.1$ and $\epsilon = 0.2$. The maximum threshold of matching distance was set as 0.33, while the value range of the matching distance was $[0, 1]$. For the three methods, the position search range was the whole $2448 \times 2050$ image. The angle search ranges were restricted to $[0^\circ], [10^\circ], [-10^\circ, 10^\circ], [-120^\circ, -60^\circ]$, and $[-10^\circ, 10^\circ]$ for the types 1, 2, 3, 4, and 5, respectively. The results of image feature extraction using the three methods were given in Fig. 10.

In comparison, the CPIE method demonstrated the best robustness, which extracted the interesting features correctly from the five types of microscopic images. The reason was that CPIE only concerns contour of interest, which was invariant with imaging condition. In addition, the time cost of the CPIE method was the smallest, as shown in Table I. The reason was that CPIE only took simple operations with a few sampling points for each transformation. Moreover, the object detection accuracies of the three methods were also shown in Table I. Here the position and angle errors were given by comparing the detected object pose and the truth object pose in the images. The CPIE method’s object detection step provided the highest accuracy. The other two methods involved many irrelevant pixels, which affected their detection accuracies.

C. Application of Geometric Constraints Using a Cylindrical Object

In this experiment, we used the cylindrical micro object to verify the effectiveness of the two geometric constraints.

1) Calibration: A micro ball was used as the calibration object. The ball’s center formed a scene point. First, the ball was fixed in the FOVs of the cameras 1 and 2.
Then each camera was translated for nine times, with the ball kept in the depth of field. A group of translation lengths and image offsets was obtained for each camera. Using the LSM, the calibrated image Jacobian matrices of the cameras 1 and 2 were obtained as 

\[
J_{c1} = \begin{bmatrix} 0.000536 & -0.000904 \end{bmatrix}^T \text{ pixel/\(\mu\)m}
\]

and 

\[
J_{c2} = \begin{bmatrix} 0.000860 & -0.003569 \end{bmatrix}^T \text{ pixel/\(\mu\)m},
\]

respectively. The parameters of the affine epipolar constraint (22) between the cameras 1 and 2 were calibrated as 

\[
a = -0.000053, b = 0.003742, c = -0.000028, d = -0.003793, e = 0.999986.
\]

The predicted best focused positions as shown in (25) for the cameras 1 and 2 were

\[
\begin{align*}
p_{c1} &= 3.42058u_{c2} - 0.006679v_{c2} - 0.018767p_{c2} + 6463, \\
p_{c2} &= -3.49897u_{c1} - 0.034913v_{c1} + 0.009047p_{c1} + 13144.
\end{align*}
\]

2) Corresponding Feature Extraction With Affine Epipolar Constraint: The object 1’s end face center produced the corresponding point features in the images of cameras 1 and 2. The epipolar lines that were determined by the corresponding image points were shown in Fig. 11(a) and (b). With the calibrated affine epipolar constraint, the position search range of CPIE was restricted to an 8-pixel-width bar region whose center was the epipolar line. Consequently, the execution time of the CPIE method on the image of camera 1 was reduced to 10.4 ms from 46.8 ms, and the execution time on the image of camera 2 was reduced to 3.4 ms from 15.4 ms. Therefore, the real time performance of corresponding image feature extraction could be improved using affine epipolar constraint, especially when the angle search range is large.

To examine the epipolar line’s accuracy, the object 1 was translated to 18 different positions. Each time, the distance \(d_e\) from the image point to its corresponding epipolar line in the image of camera 2 was recorded. Fig. 11(c) showed the distances. The root mean square (RMS) and maxima of \(d_e\) were 0.296 pixel and 0.656 pixel, respectively.

3) Autofocus With Focused Planes Intersection Constraint: The scene point to focus on was the object 1’s end face center. For the camera 1 and 2, this end face centers’ depths were the same with the object 1’s outlines from the views of the cameras 1 and 2, respectively. Using the mountain climbing strategy, the execution time of autofocusing ranged from 5 to 20 s. Then the calibrated focused planes intersection constraint was applied in the autofocusing. As long as one camera was in focus status, the other camera could take a one-time translation to the predicted best focus position, which cost only \(<1\) s.

To validate the accuracy of the focused planes intersection constraint, the object 1 was moved to 18 positions. Each time, the camera 2’s best focus position \(p_{c2}'\) was obtained using the focused planes intersection constraint. The best focus position \(p_{c2}\), which was achieved by the microscopy expert’s adjustment, was considered as the truth-value. The series of \(p_{c2}', p_{c2}\), and the error \(e_{pc} = p_{c2}' - p_{c2}\) are shown in Fig. 12. The RMS and maximum absolute value of \(e_{pc}\) was 53.4 \(\mu\)m and 113.8 \(\mu\)m, respectively, which were 12.4% and 26.5% of the 430 \(\mu\)m depth of field, respectively. Therefore, the focused planes intersection constraint gave best focus positions with the acceptable accuracy.

### Table I

<table>
<thead>
<tr>
<th>Type</th>
<th>Execution Time (ms)</th>
<th>Position Error/ Angle Error (pixel/ degree)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPIE</td>
<td>SBM</td>
<td>Fast-Match</td>
</tr>
<tr>
<td>1</td>
<td>30.7</td>
<td>60.2</td>
</tr>
<tr>
<td>2</td>
<td>40.1</td>
<td>114.8</td>
</tr>
<tr>
<td>3</td>
<td>55.2</td>
<td>284.0</td>
</tr>
<tr>
<td>4</td>
<td>548.2</td>
<td>1307.5</td>
</tr>
<tr>
<td>5</td>
<td>129.1</td>
<td>161.1</td>
</tr>
</tbody>
</table>

Note: “-” denotes failure or unavailable.

![Fig. 11](image1.png)

**Fig. 11.** Application of affine epipolar constraint. (a) Image of camera 1. (b) Image of camera 2. (c) Distances from image points to corresponding epipolar lines.

![Fig. 12](image2.png)

**Fig. 12.** Validation of focused planes intersection constraint. (a) Predicted best focus positions. (b) Focus position errors.

D. Pose Measurement of Cylindrical and Conical Objects

We used the cylindrical and conical micro objects to conduct the pose measurement experiments. The extraction of the point features in the object 1’s images and the line features in the object 2’s images is described in Section VI-B.

1) Calibration Results: A standard ball was used as the calibration object. The ball was translated to nine different positions in the world coordinates by the manipulator 1. The manipulator 1’s active motion, the cameras’ best focus positions and the image points were recorded. Using the LSM, the calibrated image Jacobian matrices of the...
cameras 1 and 2 were
\begin{align*}
J_1 &= \begin{bmatrix} -0.289783 & 0.000795 & 0.000111 & -0.289425 \\ -0.000406 & -0.000730 & -0.001342 & 0.003692 \\ 0.001277 & -0.285900 & 0.001408 & -0.285551 \end{bmatrix} \text{pixel/μm} \\
J_2 &= \begin{bmatrix} -0.007300 & 0.999947 & 0.000795 & -0.001175 \\ -0.285700 & 0.000111 & -0.000730 & 0.001277 \\ 0.001408 & -0.285900 & 0.000795 & -0.000730 \end{bmatrix} \text{pixel/μm.}
\end{align*}

The 2-DOF rotation \( R(\Delta \phi_1, \Delta \phi_2) \) of the manipulator 2 equals the combination of two rotations around the joint axis 2 and 1 sequentially, namely, \( R(\Delta \phi_1, \Delta \phi_2) = R_2(\Delta \phi_2)R_1(\Delta \phi_1) \). \( \Delta \phi_i, \nu_i \) and \( R_i \) are the rotation angle, axis direction vector, and around-axis rotation matrix of the manipulator 2’s joint \( i \) (\( i = 1, 2 \)). We fixed a stick object on the manipulator 2’s end. The manipulator 2 rotated its joint 1 to five angles when the joint 2’s angle \( \phi_2 \) was zero. Then the manipulator 2 rotated its joint 2 to five angles when the joint 1’s angle \( \phi_1 \) was zero. Each time, the stick axis’s orientation was measured. \( \nu_2 \) was calibrated as \([0.007222, 0.999947, 0.0007378]^T \) when \( \phi_1 = 0 \). \( \nu_1 \) was calibrated as \([0.999971, -0.001175, -0.0007390]^T \) when \( \phi_2 = 0 \). This calibration result was used in the following verification of orientation measurement.

2) Three-Dimensional Vector Measurement: To verify the 3-D vector measurement method, we actively translated the object 1 using the manipulator 1 for 17 times. These translation lengths ranged from 8 to 4851 μm. Each translation of the object 1’s end face center formed a vector in 3-D Cartesian space, which was measured as \( \Delta P^\prime \). The active motion \( \Delta P \) of the manipulator 1 was considered as the truth-value. The series of \( \Delta P, \Delta P^\prime \) and the position error \( e_P = \Delta P^\prime - \Delta P \) were shown in Fig. 13. The RMS and maxima of \( ||e_P|| \) were 3.00 μm and 4.34 μm, respectively. Therefore, the 3-D vector measurement held high precision even when the two scene points exceeded the optical depth of field.

For the comparison, these 3-D vectors were also measured using the method in [22], which did not consider the cameras’ linear motions in the measurement method. The RMS and maxima of the norm of position errors rose to 6.40 μm and 11.49 μm, respectively. The reason was that the camera’s optical axis was not strictly parallel to the linear motion stage’s axis. Although the angle between these axes was small, the image offset caused by camera translation would be noticeable if the camera was translated with a long distance.

3) Orientation Measurement: To verify the orientation measurement method, we actively rotated the object 2 using the manipulator 2 for 17 times. The joint angles ranged from \(-10^\circ \) to \(10^\circ \). The object 2’s orientation was indicated by the center axis of the object 2’s cylindrical body. The initial orientation was measured as \( \nu_d = [0.007678, -0.007444, 0.999943]^T \) when the joint angles were \( \phi_1 = \phi_2 = 0 \). Each time the object 2 was rotated, the measured orientation \( \nu_d^\prime \), the current joint angles \( \phi_1 \) and \( \phi_2 \) were recorded. The orientation \( \nu_d = R(\phi_1, \phi_2)\nu_d^\prime \) was considered as the truth-value, because it was determined by the exact joint angles and calibrated joint axis directions. The orientation error \( e_\theta \) was the angle between \( \nu_d^\prime \) and \( \nu_d \). The measurement results and errors were shown in Fig. 14. The RMS and maxima of \( e_\theta \) were 0.0515° and 0.106°, respectively. Therefore, the orientation measurement method provided high precision results that are uniformly represented in the world coordinates.

VII. CONCLUSION

In this paper, a multicamera microscopic vision system is developed for high precision pose measurement in 3-D Cartesian space. The CPIE method is proposed, which obtains both object-related image features and contour sharpness from grayscale images. It allows flexible reconfiguration for novel object image. The CPIE method is with good real time performance and robustness under various imaging conditions. Vectors and orientations are measured in the world coordinates using image features and image Jacobian matrices. Cameras’ linear motions are considered so that the measurement range is expanded beyond the optical depth of field without loss of accuracy. Affine epipolar constraint is utilized to increase the speed of corresponding image feature extraction. Focused planes intersection constraint is applied, which significantly reduces the time cost on multicamera autofocus.

Future work will concentrate on flexible reconfiguration of control strategy for novel precision assembly task.
REFERENCES


Fangbo Qin received the B.Sc. degree in automation from Beijing Jiaotong University, Beijing, China, in 2013. He is currently pursuing the Ph.D. degree in control science and engineering with the Research Center of Precision Sensing and Control, Institute of Automation, Chinese Academy of Sciences and the University of Chinese Academy of Sciences, Beijing.

His current research interests include robot vision and robot learning.

Fei Shen received the B.Sc. degree from Xidian University, Xi’an, China, in 2007, and the M.Sc. degree from the Beijing Institute of Technology, Beijing, China, in 2009, and the Ph.D. degree in control science and engineering from the Institute of Automation, Chinese Academy of Sciences (IACAS), Beijing, in 2012. He is currently an Associate Professor with the Research Center of Precision Sensing and Control, IACAS. His current research interests include robot control, robot vision control, and micro-assembly.

Dapeng Zhang received the B.Sc. and M.Sc. degrees from the Hebei University of Technology, Tianjin, China, in 2003 and 2006, respectively, and the Ph.D. degree from the Beijing University of Aeronautics and Astronautics, Beijing, China, in 2011. He is currently an Associate Professor with the Research Center of Precision Sensing and Control, Institute of Automation, Chinese Academy of Sciences, Beijing. His current research interests include robotics and automation, such as medical surgical robot and virtual robotic surgery.

Xilong Liu received the B.Sc. degree from Beijing Jiaotong University, Beijing, China, in 2009, and the Ph.D. degree in control theory and control engineering from the Institute of Automation, Chinese Academy of Sciences (IACAS), Beijing, in 2014. He is currently an Associate Professor with the Research Center of Precision Sensing and Control, IACAS. His current research interests include image processing, pattern recognition, visual measurement, and visual scene cognition.

De Xu (M’05–SM’09) received the B.Sc. and M.Sc. degrees in control science and engineering from the Shandong University of Technology, Jinan, China, in 1985 and 1990, respectively, and the Ph.D. degree in control science and engineering from Zhejiang University, Hangzhou, China, in 2001. He has been with the Institute of Automation, Chinese Academy of Sciences, Beijing, China, since 2001, where he is currently a Professor with the Research Center of Precision Sensing and Control. His current research interests include robotics and automation, such as visual measurement and control, microscopic vision, and micro-assembly.