TRINOCULAR VISION SYSTEM FOR POSE DETERMINATION

Mohammed A. Isa[1], Mojtaba A. Khanesar[1], Richard Leach[1], David Branson[1], and Samanta Piano[1]
[1]Manufacturing Metrology Team, Faculty of Engineering
University of Nottingham,
Nottingham, UK

POSE DETERMINATION

In modern manufacturing industries, accurate in-process pose determination for components, such as fixtures, manipulator arms and fabricated parts, is required for autonomous operations to enhance efficiency and safety in production lines [1]. The requirement to measure multiple degree of freedom feature positions makes pose measurement systems complex. As a result, three-dimensional (3D) measurement systems face operational complexity when used for pose determination [2]. The accuracy and reliability of the determined pose can be negatively affected by this complexity because factors that affect accuracy in each degree of freedom influence the pose determination collectively. Therefore, improving the performance of pose determination necessitates enhancement of the quality of all dimensional measurements on which the pose relies. Furthermore, the position estimations of all features used for pose determination should be improved and erroneous features should be minimised. This paper studies techniques for a vision-based pose determination approach and assesses the performance of the approach by comparison with a commercial measurement system—popularly known to provide more accurate measurements using interferometry [3].

Common strategies employed for pose determination include laser tracking and binocular vision [4]. Laser trackers are often more expensive and accurate when compared with vision-based systems. In traditional laser trackers, primary dimensional measurement technologies, such as interferometric or time-of-flight strategies, are integrated with angular guidance systems to allow 3D measurement of target coordinates [5]. For pose determination, further complication arises due to the need for a minimum of three 3D coordinate measurements to determine the pose of an object [6]. Vision-based 3D coordinate measurement systems are not fundamentally different from laser trackers, as both systems measure coordinate positions of targets through received light detected from the targets. However, when measurement of multiple targets is considered for pose determination, vision-based systems have the advantage of two-dimensional (2D) array sensing. The 2D sensors in vision systems allow measurement of multiple targets in relatively wide fields of view. Hence, besides the lower cost of vision systems compared to laser trackers, there is the added benefit of measuring multiple targets without further complexity in hardware. Nowadays, even commercial pose-measuring laser trackers include vision sensors to complement 3D measurement [7].

While laser trackers use dedicated retroreflectors as targets, vision systems can use arbitrary features in images as targets. However, to ensure acceptable detection and matching quality, a structured artefact with well-known dimensions is often used. Due to the prevalent algorithms in computer and machine vision literature aimed at unstructured targets, vision systems lack adequate metrological tools to assess the quality of vision-based measurements. Consequently, this paper presents the use of artefacts composed of pre-calibrated sphere features to maintain good visibility and high quality of measurements. As an established element for referencing and calibration of traditional metrological devices and tools [8], sphere can play an important role in improving the quality of vision-based measurements.

This paper proposes extending a binocular vision to a trinocular vision system for pose determination using a sphere-based artefact introduced in [9]. Binocular vision algorithms have been developed for localisation of spherical features with evaluation of the associated uncertainty [9]. Introduction of trinocular constraints can improve measurement accuracy and eliminate disagreements between multiple binocular results. Traditionally, a trinocular vision setup is treated as a set of three binocular vision systems where three measurement results are obtained and fused [10]. Due to the fact that there

is no trinocular relationship that constrains features in the three binocular vision systems, this trinocular system is referred to as multi-view stereo vision throughout this paper. Relationships between image features in binocular vision is characterised by a fundamental matrix while trinocular features are related through a trifocal tensor. Using unstructured features, no significant advantage was demonstrated in replacing the fundamental matrix with the trifocal tensor [11]. However, using structured linear features, it was shown that using the trifocal tensor improves measurement accuracy [10]. In both works, using structured and unstructured features, the measurements from the vision systems were compared against other vision measurement strategies that lack metrological traceability. It is important to evaluate the benefits of trinocular vision using features obtained from a calibrated artefact against a traceable reference measurement. As a result, this paper compares multi-stereo vision and trinocular vision strategies using reference measurements from a calibrated laser interferometer. In addition, the vision strategies were evaluated using a calibrated sphere-based artefact.

This paper begins with discussion of the basic principle of our vision-based system for spherical features. Binocular 3D coordinate measurements are briefly explained and the benefits of trinocular systems are discussed. Next, the process of trinocular characterisation is outlined and the characterisation results are compared with that from a binocular system. Finally, by the measuring coordinates of a sphere-based pose measurement artefact, the performances of the trinocular and binocular vision systems are assessed and compared.

BINOCULAR TO TRINOCULAR VISION SYSTEM

The 3D coordinates of a point can be determined by triangulation using at least two cameras or one camera and a known light projection. Binocular vision systems have been used to improve positional accuracy, enhance smart product manipulation and carry out quality inspection [1]. The overall performance of vision systems is significantly affected by how well the algorithms can recognise, match and localise features from images. To do this, binocular vision systems apply epipolar constraint to determine the projective compliance of the evaluated image feature pairs. For a binocular system with known internal projective properties, once correspondence among candidate image feature points is determined, the task of constructing the 3D points is straightforward. Thus, establishing correspondence between image feature points from two pre-calibrated cameras is analogous to finding the 3D points that resulted in the projected image points [12].

When a sphere is observed from camera views A and B, as shown in FIGURE 1, the projections of the tangential edges of the sphere are what border the observed blobs in the images. The perceived projections are elliptical and serve as the observations from which the spherical centre $\boldsymbol{S}$ and diameter are determined. Three or more spheres measured on a complex artefact can thus be used to determine the pose of the artefact. In addition, the dispersion of the observed image data can be used to improve pose determination and evaluate the associated uncertainty [9,13].

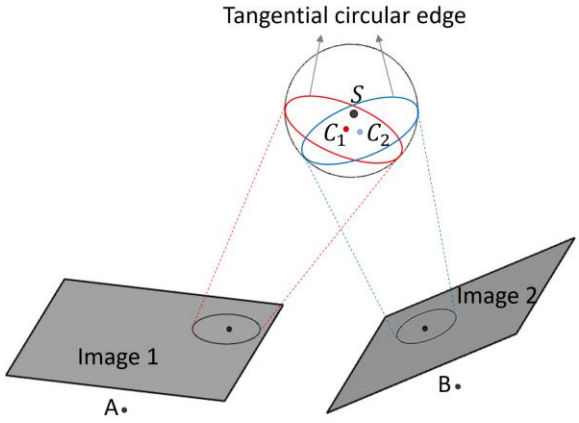

FIGURE 1. Spherical tangential circles projected as ellipses onto image 1 and 2 from views A and B respectively.

Binocular vision can produce erroneous results because the epipolar constraint for quantifying the degree of image point correspondence can fail. This is because a point in one image can match with any candidate point on the corresponding epipolar line in a second image. Adding a third camera and using image triplets for correspondence is one method to minimise epipolar constraint failure.

Theoretically, a tri-camera relationship based on a trifocal tensor, rather than three fundamental matrices, should provide a more compact trinocular relationship to improve accuracy [14].

However, the impact of trifocal tensor application on measurement accuracy has not been shown practically in many applications because the algorithms necessary for its characterisation and implementation are not readily available [11]. In the next section, the required computational steps to transition from binocular to trinocular sphere-based vision systems are discussed.

TRINOCULAR VISION CHARACTERISATION

Just as the fundamental matrix represents the projective geometric properties between two views [9], the trifocal tensor encapsulates the internal relationship in a three-view system [14]. The trifocal tensor does not depend on scene structure, thus providing the explicit relationship between observed image points in three views without any scene information. In addition to two-view correspondences that are practically only known to a degree of certainty, having a third view enforces a stricter constraint, thereby ensuring more reliable image feature triplets are obtained. The trinocular constraint is not only beneficial in enhancing image feature correspondences, it is also beneficial in triangulation of the feature point triplets to obtain more accurate 3D points.

The image projections ($\boldsymbol{x}_A$, $\boldsymbol{x}_B$ and $\boldsymbol{x}_C$) of a 3D point $\boldsymbol{X}$ observed in three camera views can be expressed through the respective mappings $\boldsymbol{P}_A$, $\boldsymbol{P}_B$ and $\boldsymbol{P}_C$ in the expressions

$$\left.\begin{aligned}\boldsymbol{x}_A &= \boldsymbol{P}_A\boldsymbol{X}\\ \boldsymbol{x}_B &= \boldsymbol{P}_B\boldsymbol{X}\\ \boldsymbol{x}_C &= \boldsymbol{P}_C\boldsymbol{X}\end{aligned}\right\}. \tag{1}$$

Expressed in homogenous coordinates, the projective mappings, commonly referred to as camera matrices, depend on the intrinsic properties and relative poses of the cameras. Before cameras can be used as measurement devices, their intrinsic properties and poses must be characterised.

The trifocal relationship, in tensor notation, for the three corresponding image points $\boldsymbol{x}_A$, $\boldsymbol{x}_B$ and $\boldsymbol{x}_C$ is given by

$$x_A^i x_B^j x_C^k \epsilon_{jqs}\epsilon_{krt}\mathrm{T}_i^{qr} = 0_{st}, \tag{2}$$

where ϵ_{ijk} represents the Levi-Civita tensor, and the indices i, j, k, q, r, s and t take the values $\{1,2,3\}$ corresponding to values of the terms. The tensor T_i^{qr} is the trifocal tensor and is the unknown during trifocal vision characterisation. During characterisation of vision systems, an artefact with easily detectable features and known feature dimensions is used to obtain reference 3D points in a 3D world coordinate system. When the features are detected in the images, the unknown terms in Equation (1) are the camera matrices. Given a set of seven or more image-point correspondences [14], the trifocal tensor can be estimated using Equation (2). However, the solution is not guaranteed to be geometrically valid [14]. The tensor needs to be constrained to obtain a geometrically relevant solution. The necessary constraint on T_i^{qr} is imposed by estimating the tensor that is compliant with the camera matrices $\boldsymbol{P}_A = [I_{3\times3} \mid 0_{3\times1}]$, $\boldsymbol{P}_B = \left[a_i^q \mid e_B^q\right]$ and $\boldsymbol{P}_C = [b_i^r \mid e_C^r]$. $\boldsymbol{P}_A$ is a canonical mapping, where $\boldsymbol{I}_{3\times3}$ and $\boldsymbol{0}_{3\times1}$ are the identity matrix and zero vector respectively. The elements e_B^q and e_C^r are the epipoles in views B and C respectively, when paired with view A. Additional properties of the trifocal tensor can be found in [14]. The trifocal tensor is related to the camera matrices by

$$\mathrm{T}_i^{qr} = a_i^q e_C^r - e_B^q b_i^r. \tag{3}$$

A 200 mm × 150 mm checkerboard artefact, manufactured by a lithographic printing process with a tolerance of 20 μm, was used for the characterisation in this paper. The three cameras were initially characterised as three binocular pairs using the MATLAB machine vision toolbox. The obtained binocular parameters were then used to obtain initial estimates for the solution of the trifocal tensor satisfying Equations (2) and (3). For the trinocular vision, the number of radial distortion parameters were expanded to six from the three supported by the MATLAB algorithms. A maximum likelihood problem (see Algorithm 16.3 in [14]) was solved for the internal trinocular parameters by first triangulating and reprojecting reconstructed 3D positions of the checkerboard points. The maximum likelihood solution was obtained by the iterative minimisation of the reprojection errors evaluated using Equation (1). FIGURE 2 compares the reprojection errors for the binocular and trinocular models. The reprojection error is significantly reduced when the trinocular model is used. This reduction indicates a more stable and accurate estimation of the internal parameters for the vision system.

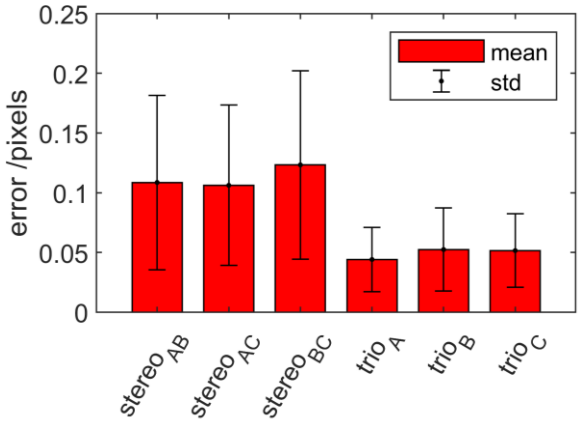

FIGURE 2. Mean and standard deviation (std) of reprojection errors from binocular view pairs ($stereo_{AB}$, $stereo_{AC}$, and $stereo_{BC}$) and trinocular reprojections in views A, B and C (given as $trio_A$, $trio_B$ and $trio_C$ respectively), evaluated from 88 checkerboard points measured at 27 checkerboard positions.

MEASUREMENT ASSESSMENT

The characterised trinocular vision system was used for pose determination from 3D coordinate measurements of spherical features [13]. The spherical features were designed to remain rigidly attached to a measurement artefact that is intended to be used for pose determination from 3D coordinate positions of three or more spheres. The position of the measurement artefact, shown in FIGURE 3 was measured using individual spherical tangential circles as shown in FIGURE 1. The 3D coordinate of each spherical feature was measured using the characterised trinocular and multi-stereo vision systems. For each 25 mm incremental displacement position shown, the three images from the cameras were recorded. Simultaneously, a Renishaw XL-80 laser interferometer was used to record the benchmark displacement of the artefact. The manufacturer-stated accuracy for displacements measured by the laser interferometer is 0.2 μm.

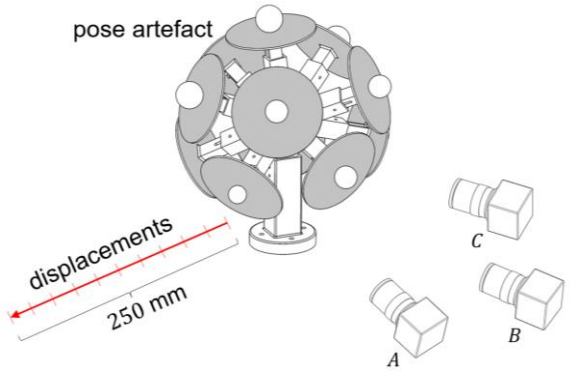

FIGURE 3. 3D coordinate measurement of pose-measuring artefact using images recorded from three camera views (A, B and C), as the artefact is incrementally displaced by 25 mm along a linear direction.

The performances of the binocular and trinocular vision systems are assessed by comparing the distances between spheres at successive displacement positions against displacements measured by the interferometer. FIGURE 4a), FIGURE 4b) and FIGURE 4c) give the displacement errors of the three binocular combinations (AB, AC, and BC respectively), while FIGURE 4d) shows the error of the complete trinocular system. The errors shown in the figure indicate the performance of the 3D coordinate measurement of the vision systems.

The three binocular 3D coordinate measurements, as shown in FIGURE 4, have varying deviations. A 3D coordinate measurement obtained by averaging or fusing the three binocular measurements can reduce random noise error [10]. This multi-stereo approach does not account for systematic error in each binocular system and is not as stable as trinocular system that constraints the internal parameters of the three-camera system to obtain a single 3D coordinate measurement. Compared to the characterisation, the 3D coordinate measurement in this section contains additional uncertainties related to sphere-feature detection process. Despite this, there are improvements in the trinocular vision measurement results shown in FIGURE 4d); the root-mean-square error (given in TABLE 1) from all the 3D coordinates of the matched displaced spheres in FIGURE 4 indicates a reduction in the error for the trinocular system. The values in the table represent the disagreement between each vision measurement with the reference interferometric displacements. The slight improvement in the trinocular

measurement shown in TABLE 1 can be expected if sphere detection errors are dominant compared to the characterisation errors.

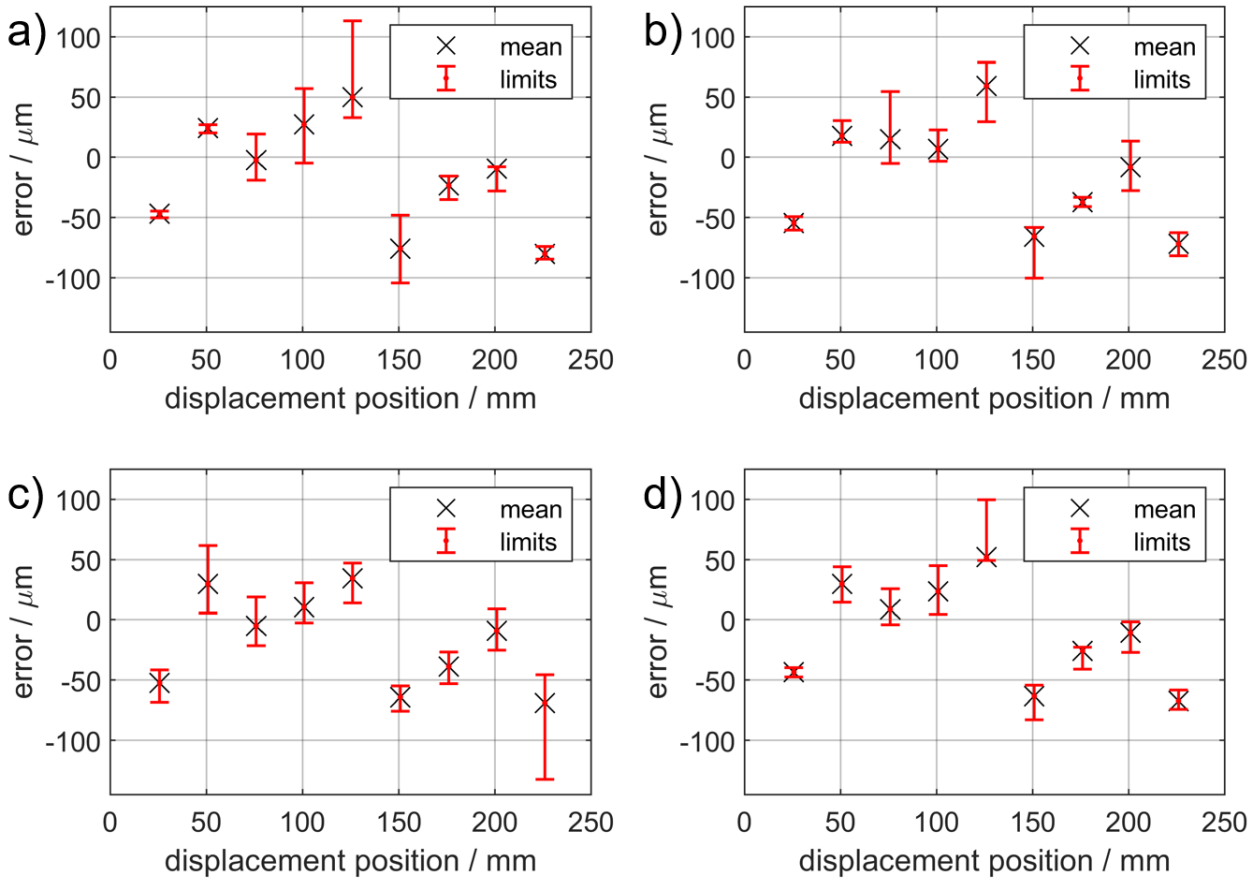

FIGURE 4. Error in 3D coordinate measurement of binocular combinations of the three camera views are given in a), b) and c). Error in the trinocular 3D coordinate measurements are given in d). Errors of vison systems are evaluated for every 25 mm displacement using laser interferometer measurement as reference. The mean, lower and upper limits indicated represent the respective mean, minimum and maximum errors from all individuals spheres step distances evaluated at each displacement position across five repeated runs.

TABLE 1. Root mean square of error in 3D coordinate measurements using binocular ($stereo_{AB}$, $stereo_{AC}$ and $stereo_{BC}$) and trinocular vision $trio_{ABC}$.

	$stereo_{AB}$	$stereo_{AC}$	$stereo_{BC}$	$trio_{ABC}$
RMS / µm	51.2	49.9	46.1	45.1

CONCLUSION

This work aims to investigate the performance of a trinocular sphere-based pose measurement system using components, algorithms and processes that prioritise metrological quality. Results from our approach are compared with binocular measurements using a reference laser interferometer. It is shown that introducing trinocular constraints to a multi-stereo vision system and increasing radial distortion

parameters can reduce reprojection errors by up to 50% and improve 3D coordinate measurements.

Future work will extend performance evaluation to individual 3D position components using a laser tracker. An improved sphere detection algorithm will also be applied and tested to further enhance the performance of the trinocular vision system.

Acknowledgements: This work is funded by the Engineering and Physical Sciences Research Council (EPSRC) under grant number: EP/T023805/1 HARISOM. The authors would like to thank UKRI Research England Development (RED) Fund for assisting this work via the Midlands Centre for Data-Drive Metrology.

REFERENCES

[1] He Z, Wu M, Zhao X, Zhang S, Tan J A Generative Feature-to-Image Robotic Vision Framework for 6D Pose Measurement of Metal Parts. IEEE/ASME Trans Mechatronics 2021:1–12.

[2] Hou D, Mei X, Huang W, Li J, Wang C, Wang X An Online and Vision-Based Method for Fixtured Pose Measurement of Nondatum Complex Component. IEEE Trans Instrum Meas 2020;69:3370–6.

[3] Leach R K Optical Measurement of Surface Topography. Springer; 2011.

[4] Guo Y, Chen G, Ma H, Ye D Researches on binocular vision pose measurement with selected feature points. Optik 2016;127:8090–5.

[5] Vikas, Sahu R K A review on application of laser tracker in precision positioning metrology of particle accelerators. Precis Eng 2021;71:232–49.

[6] Guo Z, Li Z, Zhang D, Yang L Research on pose estimation method for cooperative target based on monocular images. Proc 2011 6th IEEE Conf Ind Electron Appl ICIEA 2011:547–52.

[7] Stenz U, Hartmann J, Paffenholz J-A, Neumann I High-Precision 3D Object Capturing with Static and Kinematic Terrestrial Laser Scanning in Industrial Applications—Approaches of Quality Assessment. Remote Sens 2020;12:290.

[8] Carmignato S, De Chiffre L, Bosse H, Leach R K, Balsamo A, Estler W T Dimensional artefacts to achieve metrological traceability in advanced manufacturing. CIRP Ann 2020;00:1–24.

[9] Isa M A, Leach R, Branson D, Piano S The Effect of Motion Blur on Photogrammetric Measurements of a Robotic Moving Target. Proc. ASPE, Minnesota: 2021, 155–60.

[10] Shao M, hu M Parallel feature based calibration method for a trinocular vision sensor. Opt Express 2020;28:20573.

[11] Julià L F, Monasse P A Critical Review of the Trifocal Tensor Estimation. Image Video Technol., Wuhan: 2018, 337–49.

[12] Sims-Waterhouse D, Isa M A, Piano S, Leach R K Uncertainty model for a traceable stereo-photogrammetry system. Precis Eng 2019;63:1–9.

[13] Isa M A, Khanesar M A, Leach R, Branson D, Piano S Frequency scanning interferometry for accurate robot position measurement. EUSPEN, Geneva: 2022.

[14] Hartley R, Zisserman A Multiple View Geometry in Computer Vision. Cambridge University Press; 2003.