



Chapter 13

Measuring Full-Field Deformation in Ultra-High-Performance Concrete Structural Components Using Tag-Based Robotic Vision

Syed Zohaib Hassan, Peng “Patrick” Sun, Tiancheng Wang, Georgios Apostolakis, and Kevin Mackie

Abstract In the past few years, computer vision-based applications are widely used to determine damages in structural components. However, due to the involvement of certain levels of manual measurements, there still exists potential for improvements in the quantification of damage caused in a structure. Therefore, in this study, key-point point detection using fiducial markers is used to reduce the error in quantification of overall displacement and deformed profile of structural components (e.g., beams). This study demonstrates the use of a fiducial marker-based approach to quantify the deformation and bending profile in beam elements without physically disturbing the specimen. The visual fiducial marker (e.g., AprilTag) system has been extensively used in robotics for purposes ranging from localization of robots to increasing precision while tagging different objects in the robot’s surroundings. These AprilTag are attached to the surface of the specimen before loading. In this research, while the load is applied on the beam, the behavior of the specimen is then recorded by using a high-resolution digital camera during the experiment. Camera calibration using a checkerboard is performed to determine the extrinsic parameters (e.g., camera pose). With the help of attached AprilTags on the surface of the specimen, pixel values at the four corners and the center of each of the tags are detected during the experiment. These pixel locations are then used to get the corresponding world coordinate system by minimizing the loss function. Unlike normal bundle adjustment algorithms, camera parameters are treated as known parameters during the optimization process. Primary results show potential of proposed method in obtaining robust measurements.

Keywords Ultra-high-performance concrete · Full-field sensing · Pose estimation · AprilTag · Photogrammetry

13.1 Introduction

Structural Health Monitoring (SHM) [1] has been an active field for several decades. Monitoring the changes that a structure undergoes during its life span through adoption of modern sensors. These sensors include wireless sensors [2], or wired-sensors like Linear Variable Differential Transformers (LVDTs) [3] and improved versions of strain-gauges [4]. However, there are limitations associated with most of the present state-of-the-art SHM sensors due to their contact-based nature, there still exist chances of errors in the results. Moreover, these are only capable of providing the point data corresponding to a specific point when compared to the computer vision-based approach. In recent years, attention has been drawn to the application of computer vision for purpose of deformation measurement in structures. In this realm, the deformation is captured using the pixel locations that are extracted from camera frames [5] by fusing the vision-based data with the acceleration data. To deal with the issue of measuring the deformations from the contact-based sensors, some studies [6] utilized the application of virtual markers called imaging key points as a potential replacement of physical targets. To make the applications of machine vision-based techniques more robust, some work [7] has also been done to improve the quality of measured response of the structure in unfavorable lighting conditions by utilizing an adaptive region of interest-based image processing algorithm.

Although some of the studies discussed up till now were utilizing computer vision-based techniques to capture the deformation in a structure a small number of features were detected in a temporal manner and then converted into the final deformation or deflection values. The detection of corrected pixel values is not an easy task since most the studies rely on the

S. Z. Hassan · P. “Patrick” Sun (✉) · T. Wang · G. Apostolakis · K. Mackie
Department of Civil, Environmental, and Construction Engineering, University of Central Florida, Orlando, FL, USA
e-mail: zohaib@knights.ucf.edu; peng.sun@ucf.edu; tiancheng.wang@Knights.ucf.edu; Georgios.Apostolakis@ucf.edu; Kevin.Mackie@ucf.edu

application of virtual markers. There has been a recent development in the machine vision realm that involves non-contact-based virtual fiducial system (AprilTag) that can be automatically detected and provides a robust pixel value against any point of interest in the image frame. Some of the recent applications of fiducial markers in camera pose estimation are discussed in this work. Li [8] proposed a method of fusion the information from RGB-D images to get the robust measurements for the camera pose with help of fiducial markers. The study used RGB-D images captured by the combination of an RGB camera along with the depth capturing sensor. They concluded the improvement in translation but the error in rotation could not be minimized. Authors in [9] proposed a method of averaging the rotations captured in the image data and then the results were optimized that reduced the error in camera positioning. The proposed method was compared with other methods. Some researchers in [10] had developed an approach to build a fiducial map based on manifold optimization. These fiducial maps were then used for pose estimation and the accuracy of the results was tested against the ground truth from a motion capture system. The RMSE was found to be 3.1 cm. In regard to 3D mapping of fiducial markers, [11] studied a framework called "AprilTag3D". AprilTag3D is a setup of at least 2 tags that were not laying on the same plan but were connected with each other along an axis. In the proposed methodology, the researchers used RGB images and a third dimension to marker detector to improve Apriltag pose estimation. The proposed method improved the detection and pose estimation, especially in highly reflective environments. Gupta and Hebbalaguppe [12] presented a do-it-yourself wearable device coupled with a webcam that can provide 6 DOF per fingertip tracking in real-time. The authors used ViSP, which had the AprilTag module built-in. Abbas et al. [13] proposed a three-step pose estimation. These steps include (1) correction in trigonometric correction of yaw angle, (2) use of custom made gimbal for the improvements in yaw angle, and (3) utilizing a Gaussian process-based regression of experimental data a pose-indexed probabilistic sensor error model of AprilTag was presented and the validation was particle filter tracking.

Up till now, the applications of robotic vision in pose estimation of different objects have been discussed. Since the focus of this research is to deploy and use effectively the fiducial markers (a type of computer vision-based technology) in the domain of Civil Engineering (specifically in Structural Engineering). Therefore, a brief overview of current developments of these computer-vision-based techniques in the domain of Civil Engineering is presented here. Starting from Valença and Carmo [14] in which the researchers had proposed a method based on photogrammetry and image processing known as "Photo-Node". This one type of non-contact method was used to monitor the concrete surfaces and to measure the relevant physical parameters of reinforced concrete beam-column joints during experimental testing phase. The proposed method was used to exploit post-processing features to evaluate joint distortion, principle directions, joint rotation, and stress along with developed crack patterns and load paths. In another attempt to use computer vision-based technique to monitor the damage in a structure, the authors of Saravanan and Nishio [15] had carried out experiments using low-cost cameras and 2D digital image correlation (2D-DIC). The quasi-static loading was applied to the specimens (aluminum plate girders bridge model) in an indoor environment in a bending test and the in-plane strain field variations near the endpoint of girders were estimated in different boundary conditions. Experimental strains obtained in the tests were then compared with the strains resulting from 2D-DIC. The authors also suggested extending the study by comparing the results with the finite element analysis (FE analysis) as an extension of the study. In an attempt to measure the structural deformation using one of the most feasible approaches, researchers in Kromanis et al. [16] proposed a study that examined the performance of the structure by monitoring the structural deformation by using the video/images captured from smartphones when the structure (i.e., beams) was subjected to the static, dynamic and as well as quasi-static loading in a laboratory environment. The conclusions were made based on the RMSE of deviation between the results of image analysis methods and the measurement collections. In Xiang et al. [17] an approach focused on information shared between the cameras is used to increase the robustness of the previously developed algorithm. The authors of the study emphasized the possibility of a camera position-independent imaging approach of the condition assessments of structures focusing on the fact that the structural features remain the same between the cameras even if the images are captured from different angles of camera positions. The DeforMonit application technique is developed at Nottingham Trent University by R Kromanis and A Al-Habaibeh [18] to evaluate target displacements.

The main objective of this study is to present a framework to estimate the pose of a 3D object in the real world by deploying a multi-camera model approach. The robustness in the algorithm is provided by the bundle adjustment method that is used for the scene reconstruction from multi-view cameras. In this paper, the first section has provided a broad horizon of the current state-of-the-art of using machine vision and fiducial markers in structural health monitoring purposes. The second section will discuss the methodology that is used to implement the proposed techniques, the verification of the proposed methodology is then provided in experimental section in which the experimental layout will be discussed followed by the results and discussion part where the analysis and conclusions will be drawn and potential future works will also be discussed in the future work section.

13.2 Methodology

13.2.1 Pinhole Camera Model

In computer vision-based studies, understating camera models and their characteristics is of great importance. In order to understand the projection of the 3D world coordinate system (WCS) into 2D sensor coordinate system (SCS), the pine-hole camera model is discussed in this section.

As shown in Fig. 13.1a, a light beam illuminated by a point (\mathcal{P}) in 3D WCS can be projected into 2D space of image coordinate system (ICS) and sensor coordinate system (SCS). As the first stage, translation and rotation are performed for coordinates of any point \mathcal{P} (${}^w\mathbf{X}_P$) to transform it from WCS to camera coordinate system (CCS) (${}^c\mathbf{X}_P$) as shown in Eq. 13.1. At the further stages, the point is projected onto 2D image coordinate system (ICS) (denoted as ${}^i\mathbf{x}_p$ or $({}^i x_p, {}^i y_p)$) and the sensor coordinate system (SCS) (denoted as ${}^s\mathbf{u}_p$ or $({}^s u_p, {}^s v_p)$).

The transformation from WCS to CCS is expressed as:

$$\begin{bmatrix} {}^c X \\ {}^c Y \\ {}^c Z \\ 1 \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} & 0 \\ r_{21} & r_{22} & r_{23} & 0 \\ r_{31} & r_{32} & r_{33} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & -{}^w X_C \\ 0 & 1 & 0 & -{}^w Y_C \\ 0 & 0 & 1 & -{}^w Z_C \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} {}^w X \\ {}^w Y \\ {}^w Z \\ 1 \end{bmatrix} \quad (13.1)$$

where $({}^w X_C, {}^w Y_C, {}^w Z_C)$ is the location of camera origin (pinhole camera) in WCS.

The equation above is the same as:

$${}^c\mathbf{X}_P = \begin{bmatrix} {}^c\mathbf{R} & -{}^c\mathbf{R}{}^w\mathbf{X}_C \\ \mathbf{0}^T & 1 \end{bmatrix} {}^w\mathbf{X}_P \quad (13.2)$$

where ${}^c\mathbf{R} \in \mathbb{R}^{3 \times 3}$; the homogeneous coordinate of point \mathcal{P} (${}^w\mathbf{X}_P$) is represented by its Euclidean coordinate $({}^w\mathbf{X}_P)$: ${}^w\mathbf{X}_P = [{}^w\mathbf{X}_P, 1]^T$. $\mathbf{X}_P \in \mathbb{R}^3$, $X_P \in \mathbb{R}^3$

The Euclidean (rigid) matrix, ${}^c_w\mathbf{H}$, is defined as:

$${}^c_w\mathbf{H} = \begin{bmatrix} {}^c\mathbf{R} & -{}^c\mathbf{R}{}^w\mathbf{X}_C \\ \mathbf{0}^T & 1 \end{bmatrix}$$

The transformation from CCS to ICS can be expressed as:

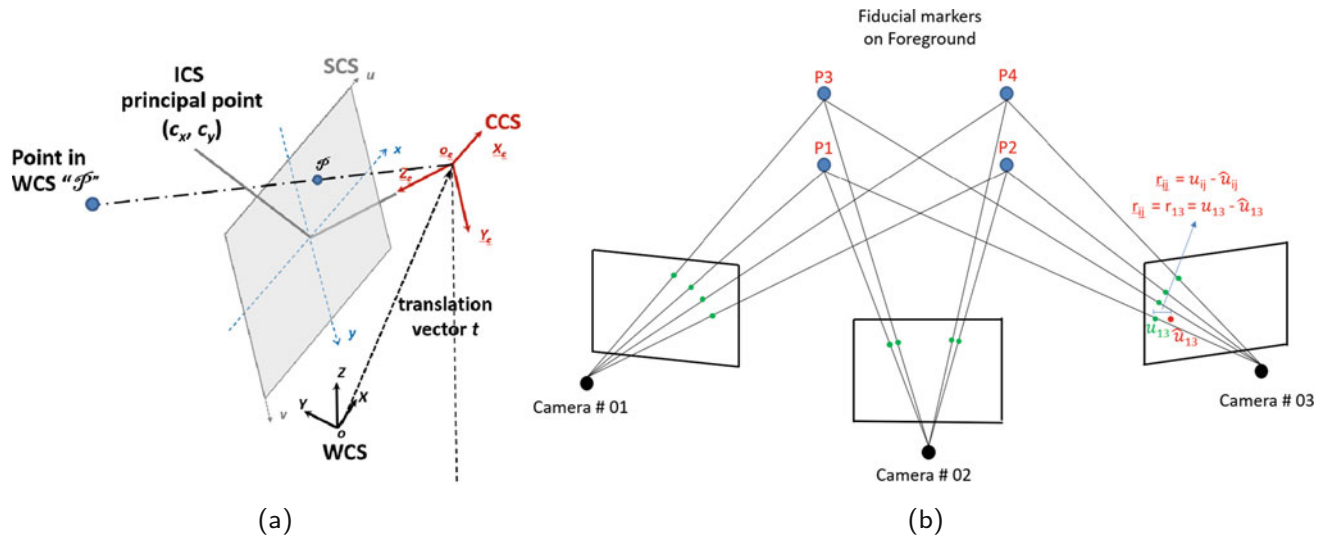


Fig. 13.1 Schematics of (a) projection of a point (P) in 3D world coordinate system (WCS) to p in 2D sensor coordinate system (SCS), (b) three cameras setup focusing on four points

$$\begin{bmatrix} iu_p \\ iv_p \\ iw_p \end{bmatrix} = \begin{bmatrix} f_x & 0 & 0 & 0 \\ 0 & f_y & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} cX_P \\ cY_P \\ cX_P \\ 1 \end{bmatrix} \quad (13.3)$$

where $ix_p = iu_p/iw_p$ and $iy_p = iv_p/iw_p$ so as to convert homogeneous coordinates (i.e., $[iu_p, iv_p, iw_p]^T$) to Euclidean coordinates (i.e., $[ix_p, iy_p]^T$).

The projection matrix, \mathbf{M}_{proj} , is defined as:

$$\mathbf{M}_{\text{proj}} = \begin{bmatrix} f_x & 0 & 0 & 0 \\ 0 & f_y & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

The transformation from ICS to SCS is expressed as:

$$\begin{bmatrix} sx_p \\ sy_p \\ 1 \end{bmatrix} = \begin{bmatrix} S_x & S_{xy} & s_{xH} \\ 0 & S_y & s_{yH} \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} ix_p \\ iy_p \\ 1 \end{bmatrix} \quad (13.4)$$

The affine matrix, $\mathbf{M}_{\text{affine}}$, is defined as:

$$\mathbf{M}_{\text{affine}} = \begin{bmatrix} S_x & S_{xy} & s_{xH} \\ 0 & S_y & s_{yH} \\ 0 & 0 & 1 \end{bmatrix}$$

Hence, the overall projection from WCS to SCS can be expressed as:

$$\begin{bmatrix} su_p \\ sv_p \\ sw_p \end{bmatrix} = \mathbf{M}_{\text{affine}} \mathbf{M}_{\text{proj}} \mathbf{H} \begin{bmatrix} wX_P \\ wY_P \\ wZ_P \\ 1 \end{bmatrix}$$

where $\begin{bmatrix} sx_p \\ sy_p \end{bmatrix} \sim \begin{bmatrix} su_p \\ sv_p \\ sw_p \end{bmatrix}$ which means $sx_p = su_p/sw_p$ and $sy_p = sv_p/sw_p$.

13.2.2 Bundle Adjustment Algorithm

The 3D reconstruction of a point from 2D in 3D has been an active direction within the photogrammetry realm. There are two main frameworks as described in the work [19]. One framework is referred to as ‘‘Structure from motion’’ (SfM) that is a unifying framework that works on the principle of estimating the motion and the structure simultaneously, while the other framework is referred to as decoupling framework, also known as motion estimation where the estimation of structure and motion are performed separately. To work with the decoupling framework a prototypical approach based on epipolar-constrained is utilized. In this approach, the space of essential matrices is used as the optimization criterion where the deviation is due to the measurements errors (especially the noise in images) [20]. This type of approach has a prominent benefit on unifying framework method since it works on a low-dimensional search space. However, the drawbacks of this approach include statistically biased estimation of translation [21, 22] and its sensitivity to input images noise that indicates the degraded performance of decoupling framework [23]. On the contrary, for the unifying framework, one of the most popular approaches is known as the Bundle Adjustment. This optimization approach minimizes the error between the predicted and observed pixel values.

Bundle Adjustment is used to minimize the sum of errors with the observed 2D pixel values and the measured 2D pixel values where the measured pixel are obtained as a projection function of predicted 3D points by using camera parameters. These measured 2D pixel values are obtained by using the camera model approach as described in detail in Sect. 13.2.1. In other words, the Bundle Adjustment technique is used to minimize the error between detected and projected 2D images pixels using the parameters from multiple camera perspectives. Using Fig. 13.1b, the Bundle Adjustment method can be represented as a nonlinear least square problem that can be given as:

$$\min_{C, X} \sum_{i=1}^n \sum_{j=1}^m (u_{ij} - \pi(C_j, X_i))^2 \quad (13.5)$$

where u_{ij} is the observation in SCS corresponding to the i -th 3D point with coordinates ${}^w X_i$ observed by the j -th camera C_j . Let r_{ij} be the vector:

$$r_{ij} = (u_{ij} - \pi(C_j, X_i)) \quad (13.6)$$

where $r_{ij} \in \mathbb{R}^{1 \times 2}$, The points " u_{ij} " and " \hat{u}_{ij} " are the observed and measured 2D pixel value of a same point from a camera. Therefore, Eq. 13.6 can be written as:

$$\min_{C, X} r^T r \quad (13.7)$$

where $r = [r_{11}, r_{12}, \dots, r_{1m}, \dots, r_{n1}, r_{n2}, \dots, r_{nm}]^T$ and $r \in \mathbb{R}^{2mn \times 1}$.

Vector r could be expanded by using Taylor's expansion as:

$$r(x + \delta x) = r(x) + g^T \delta x + \frac{1}{2} \delta x^T H \delta x \quad (13.8)$$

By taking the derivative of Taylor's expansion in Eq. 13.8 and equating it to zero, we can get:

$$H \delta x = -g \quad (13.9)$$

where H is Hessian and g is the gradient of vector r . Therefore, H can be replace by $J^T J$ and g by $J^T r$. The simplified equation can be expressed as:

$$J^T J \delta x = -J^T r \quad (13.10)$$

13.2.3 Levenberg-Marquardt Algorithm

Because the Bundle Adjustment algorithm usually deals with large error minimization problems, an iterative optimization tool is needed to reduce the global loss until it converges. The Levenberg–Marquardt (LM) algorithm is one of the popular tools. The main reason for selecting LM as here as a solution to this nonlinear problem is its ability to work as a gradient descent method and as the Gauss–Newton method. This algorithm adaptively varies parameters updates between the gradient descent method and the Gauss–Newton method. Initially, when the parameters are far away from the optimal value, it behaves as a gradient descent method; when the solution converges close to the optimal value it behaves like the Gauss–Newton method.

To understand LM algorithm, let us consider a function, f , that maps the parameter vector p and calculates the measurement pixel vector $\hat{x} = f(p)$. The parameter vectors is a multi-dimensional vectors consisting of camera rotation \mathbf{R} , translation, \mathbf{t} and 3D location (${}^w X$) in WCS. Therefore, the parameter vector is represented as $p(\mathbf{R}, \mathbf{t}, {}^w X)$

$$\hat{x} = f(\mathbf{R}, \mathbf{t}, {}^w X) \quad (13.11)$$

The purpose of LM algorithm is to find the optimal vector p^+ so that the re-projection error $\epsilon = x - \hat{x}$ is minimized.

To solve Eq. 13.10, we use the LM algorithm and the expression can be derived as,

$$(\mathbf{J}^T \mathbf{J} + \lambda \mathbf{I}) \delta \mathbf{x} = -\mathbf{J}^T \mathbf{r} \quad (13.12)$$

Equations 13.10 and 13.12 can be solved directly through algebra system of equations but for the larger system of equations, e.g., for one thousand cameras with two Million 3D points, the solution will be impractical. Therefore, for the sake of convenience the optimization problem is solved by using Eq. 13.10. The predicted pixel value is $\hat{u}_{ij} = \pi(\mathbf{C}_j, \mathbf{X}_i)$ and the parameter block $(\mathbf{C}_j, \mathbf{X}_i)$ is further categorized into two sub-block, i.e., the camera block “c” and the structure or 3D point block “p” as $\mathbf{x} = [\mathbf{c}, \mathbf{p}]$.

For the sake of simplicity, we consider an example of 3 cameras ($m = 3$) that are observing a set of 4 3D points ($n = 4$). Therefore, to construct the Jacobi matrix for this assumed condition we have $A_{ij} = \frac{\partial \hat{u}'_{ij}}{\partial c_j}$ and $B_{ij} = \frac{\partial \hat{u}'_{ij}}{\partial p_j}$ therefore, the sparse Jacobi matrix will be constructed as shown in Eq. 13.13 below.

$$\mathbf{J} = \frac{\partial \hat{\mathbf{u}}}{\partial \mathbf{x}} = \begin{bmatrix} A_{11} & 0 & 0 & B_{11} & 0 & 0 & 0 \\ 0 & A_{12} & 0 & B_{12} & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & A_{42} & 0 & 0 & 0 & 0 & B_{42} \\ 0 & 0 & A_{43} & 0 & 0 & 0 & B_{43} \end{bmatrix} \quad (13.13)$$

However, the camera calibration step is performed beforehand in the present research. Therefore, the camera parameters A_{ij} are taken as constant for all the three cameras used here in this paper. The modified Jacobian matrix will only consist of B_{ij} since the point parameters (3D WCS) are unknown.

13.3 Experimental Setup

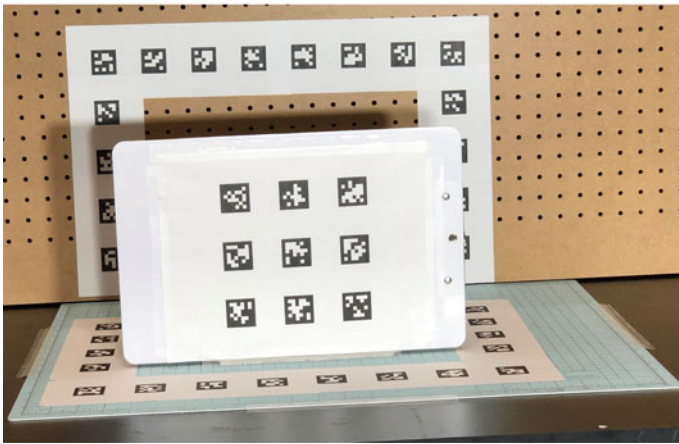
This section will provide a detailed description of the experiment that is conducted in an indoor laboratory environment to implement the methodology that is described in Sect. 13.2.2. To examine the bundle adjustment techniques for multi-perspective sensing in a controlled environment, the 3D point of an object is viewed from 3 different cameras at the same time. Please note that the cameras are calibrated by using the famous checkboard calibration technique [24] before the experiment is performed. A total of three cameras are used in this experiment as shown in Fig. 13.2a and b. Sony Alpha 6000 (camera#01) and Sony Alpha 6400 (camera#03) are deployed at the extreme left and extreme right of experimental layout, respectively. Whereas a GoPro HERO10 (camera#02) is placed in between the two Sony cameras. After the calibration, the errors in internal calibration of cameras #01, 02, and 03 are observed to be 0.054, 0.062, and 0.033 pixels respectively. Errors in external calibration of cameras #01, 02, and 03 are found to be 4.371, 8.285, and 3.200 pixels respectively.

13.3.1 Tag-Based Detection Method

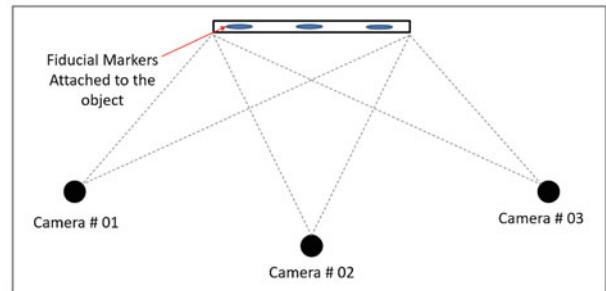
A white board attached with fiducial markers is selected as the object of interest. A total of 9 fiducial markers (i.e., AprilTag of family t36h11) are attached to the face of the object (i.e., foreground). Additionally, 32 AprilTags of family t25h9 are attached to the background (i.e., the workbench). The purpose of using separate background and foreground tag families is to achieve an undisturbed pose for each of cameras during the experiment. The center point of the work bench (i.e., the background in this case) is taken as the origin of the world coordinate system so that the position of each fiducial maker is known in WCS to serve as ground truth values. In the current experimental setup, since the orientation of the board (i.e., the foreground) is to be used for 3D reconstruction while capturing the images, it is important to keep the cameras and the background static. Hence, no motion is allowed in the camera position. A total of 32 fiducial markers (AprilTag of family t25h9 with size 30 cm) are used to identify the poses of all three cameras used in this experiment. The tags are attached to the background (static) board as can be seen in Fig. 13.2b. After all the three cameras and the background are fixed, the foreground (i.e., the board) is focalized as the region of interest is focused. A total of 9 fiducial markers (tag family of t36h11 with size 30 cm) are attached to the object of interest as foreground tags.



(a)



(b)



(c)

Fig. 13.2 Photos of (a) experimental setup for evaluation, (b) object of interest (i.e., the white board as foreground) placed on the workbench (i.e., background), and schematic view of (c) the experimental setup

13.3.2 Deployments of Cameras

The cameras are placed in an orientation such that all the 3 cameras are able to capture any movement that is performed by the object, as shown in Fig. 13.2a and c. Additionally, to improve the visual conditions, 2 LED lights are used to improve the data that is being collected during the experiment.

13.3.3 Concrete Beam Bending

As an application of the proposed technique for identification and rectification of deformation in structural members using a non-contact approach, a monocular camera setup is used to quantify the deformation of beam during destructive four-point bending test. In this experiment, the specimen constructed from Ultra-High Performance Concrete with the dimensions of $9 \times 3 \times 3$ inch ($228.6 \times 76.2 \times 76.2$ mm) is used. A total of 9 AprilTag (from the family of t25h9), evenly deployed on the face of the specimen. Sony Alpha6400 camera with a focal length of 50 mm is used to capture the data. The camera is calibrated beforehand by using checkboard calibration. Additional LED lights are deployed around the experimental area to improve the lighting conditions.

Ultra-High Performance Concrete (UHPC) is a special type of concrete. It is specially designed for extreme loading conditions because, unlike the normal concrete that has a compressive strength of around 4 to 6 ksi, UHPC is designed to achieve the strength of 18 to 22 ksi. The UHPC used in this study is designed using steel fibers (of length 7 mm) providing the necessary reinforcement along with cement and some admixtures. UHPC has the tendency to self-compact as compared to normal concrete that requires additional means of compaction. Therefore, it is referred to as self-compacting concrete. However, the properties come at a high cost. Compared to normal concrete, the UHPC costs 10 times higher. It also requires extra attention during the pouring and curing stages, for example, during the pouring process, it has to be poured as soon as possible because it hardens quicker than ordinary concrete and in the curing stage it requires a fresh round of water to be sprayed during its initial 24 hours of hardening time.

13.4 Result and Discussion

13.4.1 Tag-Based Detection

The AprilTag detection results is shown in Fig. 13.3 for the evaluation experiment. Figure 13.3(a)–(c) show detected pixel locations (red dots) and projected pixel location (blue circles) of the AprilTags attached to the background/workbench on the camera#1-3, respectively. A sample background (green square) and foreground (pink square) tags detection from camera#01 is shown in Fig. 13.3(d). The AprilTag detections results on concrete beam [shown in Fig. 13.4(a)] are performed before loading and after loading, as shown in Fig. 13.4(b)–(c), respectively. Figure 13.5a–c show the selection of randomized 3D location as an initial step to start the least square optimization algorithm (LM) for calculation of pixels (measured pixel)

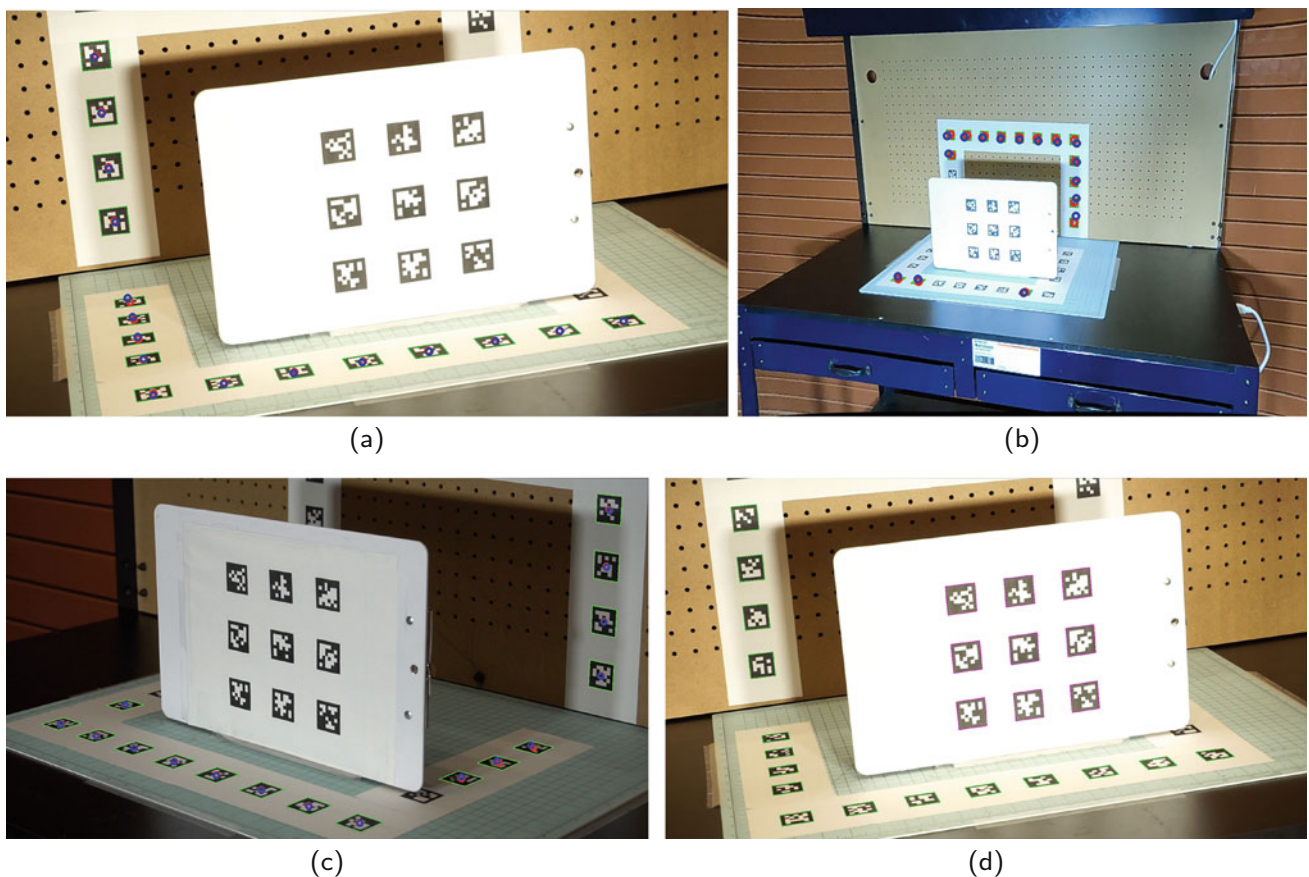
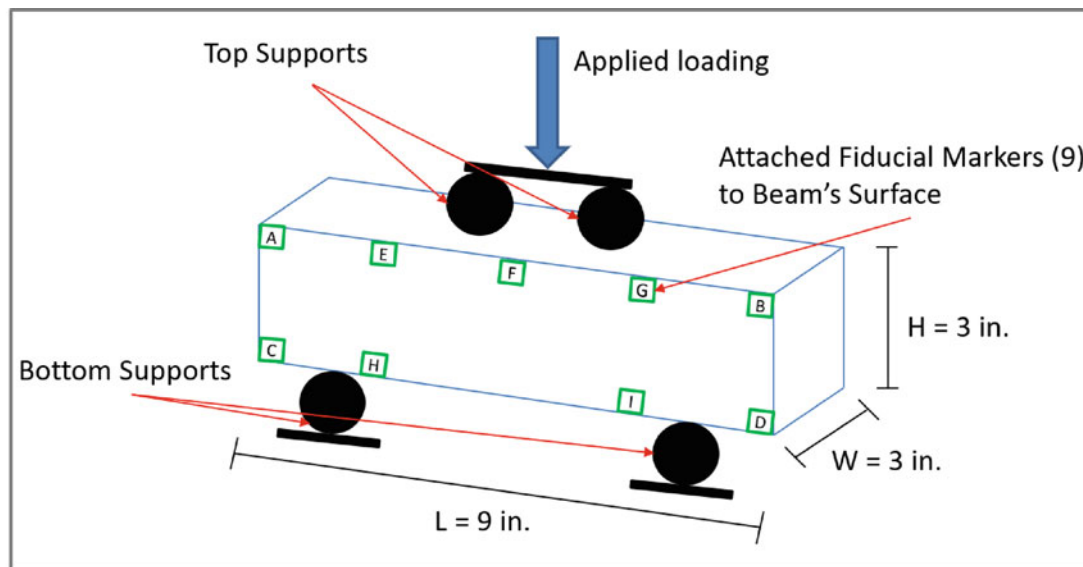
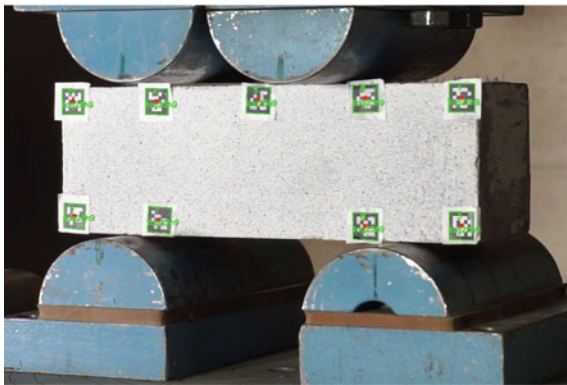


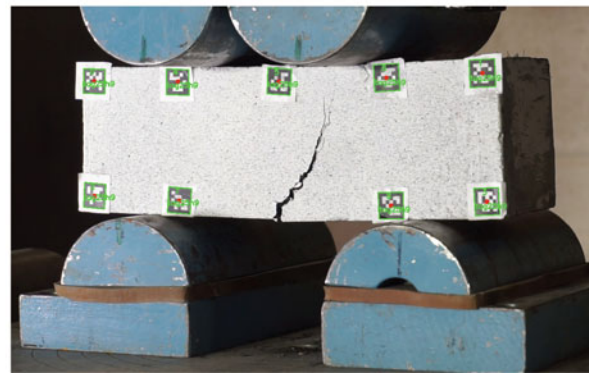
Fig. 13.3 Images of (a) detected pixel location (red dot) and projected pixel location (blue circle) on the background, (b) detected (red dot) and projected (blue circle) pixels on the background after calibrating camera#02, (c) detected (red dot) and projected (blue circle) pixels on the background after calibrating camera#03 and, (d) sample background (green square) and foreground (pink square) tags detection from camera#01



(a)



(b)



(c)

Fig. 13.4 (a) Schematic view of beam with fiducial markers attached on the surface, and images of (b) beam with tag detection before loading, (c) beam with tag detection after loading

as a function of camera parameters and 3D world locations. Results in Fig. 13.5d to f show optimized 3D location that is obtained at the end of the least square estimation. The initial error is reduced from 400 (in pixels) to about 26 pixels at an optimal solution. It is found that after the least square optimization the reconstructed world coordinates of 45 points are in good match with the provided ground truth. Hence, the proposed approach achieved an accurate estimation of structural deformation using multi-view reconstruction.

13.4.2 Concrete Beam Bending

The vertical deflection result is captured by the proposed non-contact tag detection approach, as shown in Fig. 13.6. 5 out of 9 tags that are attached to the specific location of the beam's surface are presented. Tag-0,1,2, and 3 are attached to the top left, top right, bottom left, and bottom right corner of the beam respectively, while tag#5 is attached to top center of the beam. These values are compared with the vertical deflection values provided directly by UTM as shown in Fig. 13.6a.

The results from Fig. 13.6b–d provide good agreement with that of from the UTM (i.e., closed to 6 mm). While from Fig. 13.6e, it can be anticipated that since the tag#5 is attached to the mid-point of the beam at the top, therefore, after a while in the experiment, this tag, along with the beam, is moved downward as the load is exerted from the upper loading

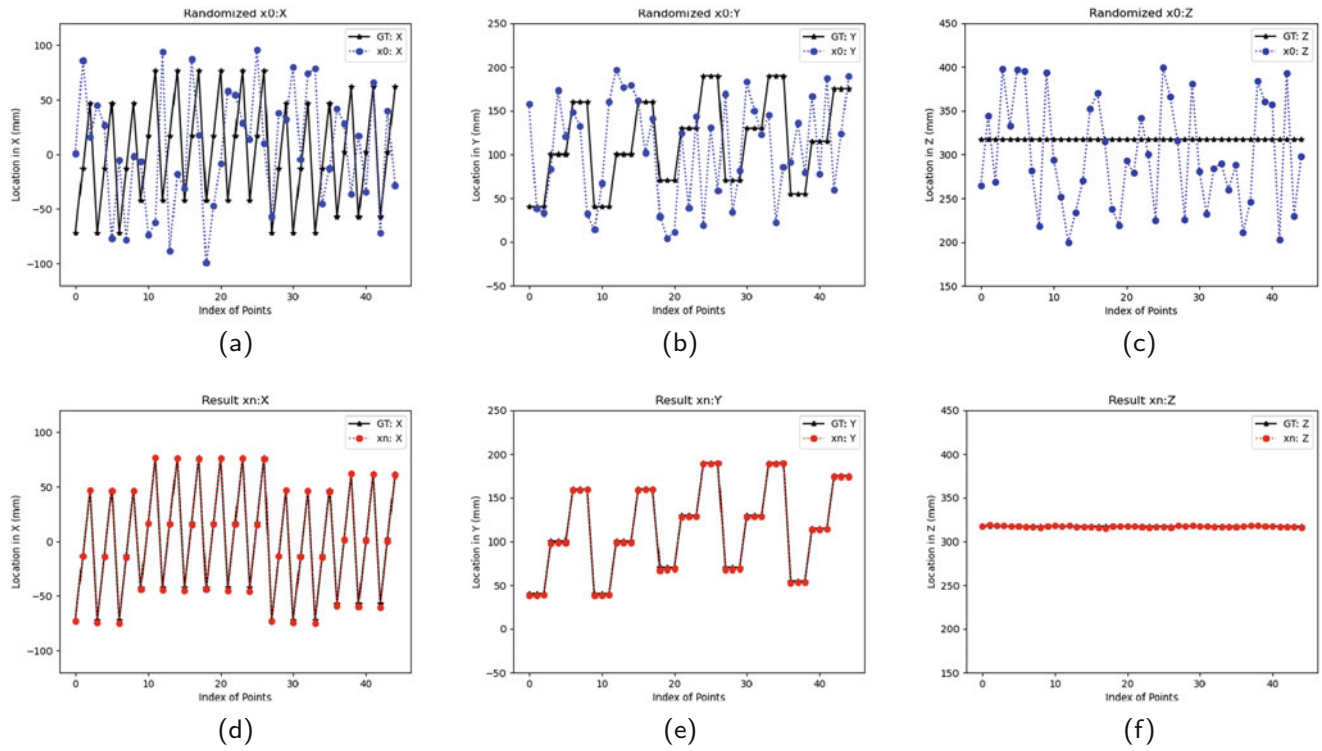


Fig. 13.5 Plots of (a), (b), (c) initialized random XYZ (3D) location (${}^w X_p^{(0)}$) to start the optimization algorithm, (d), (e), (f) optimized values in a 3D location (${}^w X_p^{(n)}$) corresponding to each point

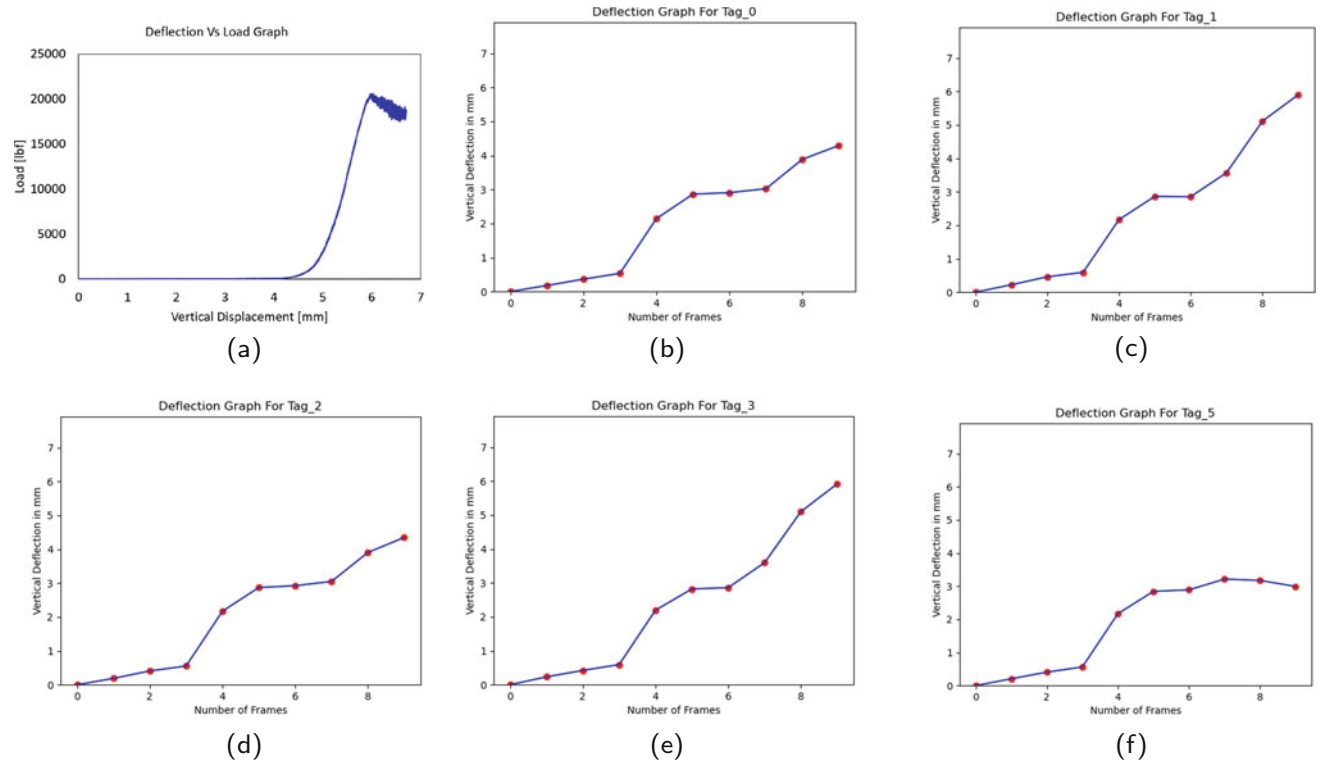


Fig. 13.6 (a) Load vs displacement from UTM, (b) Vertical deflection in Tag-A (Top left), (c) Vertical deflection in Tag-B (Top right), (d) Vertical deflection in Tag-C (Bottom left), (e) Vertical deflection in Tag-D (Bottom right), (f) Vertical deflection in Tag-F (Top center)

heading of UTM. Therefore, we can see an increase in vertical deformation at tag#5 (tag-F) that later converts into a curve and then reduces once the mid-section of the beam starts going down in vertical direction.

13.5 Conclusion and Future Work

The study aims to provide a full-field, non-contact measurement approach for a four-point bending test on UHP concrete beams using tag-based robotic vision. An evaluation experiment is carried out on a board attached with AprilTags using images taken from multiple cameras with overlapped perspectives. Shared feature points are detected using tag detection algorithms. The 3D reconstruction is treated as a Bundle Adjustment problem and the optimization is conducted by the Levenberg—Marquardt algorithm. Primary results show a good match between location estimation and ground truth. As the extension of the current work, the authors will reconstruct the 3D geometry and deformation by using multi-perspective imaging system and a more robust algorithm.

Acknowledgments The authors appreciate the support from the University of Central Florida with SEED fund (AWD00000896) and would like to thank Graduate Assistant Mr. Seyyed Sina Shid Moosavi for the help with the data analysis.

References

- Hoult, N.A., Glisic, B.: Editorial: Structural health monitoring of bridges. *Front. Built Environ.* **6**, 17 (2020)
- Geoffrine, J.M.C., Geetha, V.: Energy optimization with higher information quality for SHM application in wireless sensor networks. *IEEE Sensors J.* **19**, 3513–3520 (2019)
- Kwiecień, A., Krajewski, P., Hojdys, Ł., Tekieli, M., Słoński, M.: Flexible adhesive in composite-to-brick strengthening—experimental and numerical study. *Polymers* **10**(4), 356 (2018)
- Choi, J.-H., No, Y.-S., So, J.-P., Lee, J.M., Kim, K.-H., Hwang, M.-S., Kwon, S.-H., Park, H.-G.: A high-resolution strain-gauge nanolaser. *Nat. Commun.* **7**(1), 1–8 (2016)
- Park, J.-W., Moon, D.-S., Yoon, H., Gomez, F., Spencer Jr., B.F., Kim, J.R.: Visual–inertial displacement sensing using data fusion of vision-based displacement with acceleration. *Struct. Control Health Monit.* **25**(3), e2122 (2018) e2122 STC-17-0066.R2
- Khuc, T., Catbas, F.N.: Computer vision-based displacement and vibration monitoring without using physical target on structures. *Struct. Infrastruct. Eng.* **13**(4), 505–516 (2017)
- Lee, J., Lee, K.-C., Cho, S., Sim, S.-H.: Computer vision-based structural displacement measurement robust to light-induced image degradation for in-service bridges. *Sensors* **17**(10), 2317 (2017)
- Li, W.: A Geometry Reconstruction And Motion Tracking System Using Multiple Commodity RGB-D Cameras. PhD Thesis, The George Washington University (2020)
- Ch’ng, S.-F., Sogi, N., Purkait, P., Chin, T.-J., Fukui, K.: Resolving marker pose ambiguity by robust rotation averaging with clique constraints. In: 2020 IEEE International Conference on Robotics and Automation (ICRA), pp. 9680–9686. IEEE, Piscataway (2020)
- Hu, X., Jakob, J., Per, K., Jiang, W.: Accurate fiducial mapping for pose estimation using manifold optimization. In: 2018 International Conference on Indoor Positioning and Indoor Navigation (IPIN), pp. 206–212 (2018)
- Mateos, L.A. (2020). AprilTags 3D: Dynamic Fiducial Markers for Robust Pose Estimation in Highly Reflective Environments and Indirect Communication in Swarm Robotics. *ArXiv, abs/2001.08622*.
- Gupta, O., Hebbalaguppe, R.: FingertipCubes: An inexpensive DIY wearable for 6-DoF per fingertip pose estimation using a single RGB camera. In: SIGGRAPH Asia 2018 Posters, pp. 1–2 (2018)
- Abbas, S.M., Aslam, S., Berns, K., Muhammad, A.: Analysis and improvements in AprilTag based state estimation. *Sensors* **19**(24), 5480 (2019)
- Valença, J., Carmo, R.N.F.: Method for assessing beam column joints in RC structures using photogrammetric computer vision. *Struct. Control Health Monitor.* **24**(11), e2013 (2017)
- Saravanan, T.J., Nishio, M.: Estimation of in-plane strain field using computer vision to capture local damages in bridge structures. In: Gelman, L., Martin, N., Malcolm, A.A., (Edmund) Liew, C.K. (eds.) *Advances in Condition Monitoring and Structural Health Monitoring*, pp. 461–469, Singapore. Springer, Singapore (2021)
- Kromanis, R., Xu, Y., Lydon, D., Martinez del Rincon, J., Al-Habaibeh, A.: Measuring structural deformations in the laboratory environment using smartphones. *Front. Built Environ.* **5**, 44 (2019)
- Xiang, J., Yang, Z., Aguilar, J.: Structural health monitoring for mechanical structures using multi-sensor data. *Int. J. Distrib. Sens. Netw.* **14**, 155014771880201, 09 (2018)
- Kromanis, R., Al-Habaibeh, A.: Low cost vision-based systems using smartphones for measuring deformation in structures for condition monitoring and asset management. In: *The 8th International Conference on Structural Health Monitoring of Intelligent Infrastructure (2017)*
- Han, Y.: Relations between bundle-adjustment and epipolar-geometry-based approaches, and their applications to efficient structure from motion. *Real-Time Imag.* **10**(6), 389–402 (2004)
- Mühlich, M., Feiden, D., Mester, R.: A note on error metrics and optimization criteria in 3d vision. In: *Proceedings of the IEEE Workshop on Vision, Modeling and Visualization, Erlangen*, pp. 149–156 (1999)

21. Jepson, A.D., Heeger, D.J.: Linear subspace methods for recovering translational direction. *Spatial Vision in Humans and Robots*, pp. 39–62 (1992)
22. Kanatani, K.: 3-d interpretation of optical flow by renormalization. *Int. J. Comput. Vis.* **11**(3), 267–282 (1993)
23. McLauchlan, P.F., Murray, D.W.: A unifying framework for structure and motion recovery from image sequences. In: *Proceedings of IEEE International Conference on Computer Vision*, pp. 314–320. IEEE, Piscataway (1995)
24. Zhang, Z.: A flexible new technique for camera calibration. *IEEE Trans. Pattern Analy. Mach. Intell.* **22**, 1330–1334, 12 (2000)