DOI: 10.3217/978-3-99161-057-1-051

# Development of a Wireless Stereo Vision System for 3D Displacement Online Long-Term Monitoring of Tall Structures

Miaomin Wang, Ki-Young Koo
Vibration Engineering Section, Faculty of Environment, Science and Economy, University of Exeter, Exeter, EX4 4QF, UK

ABSTRACT: This paper presents a novel wireless stereo vision system for 3D displacement monitoring of tall structures. The system uses two GNSS wireless camera nodes to capture images of a target and calculate its 2D displacement independently. Each node uploads its data to the cloud, where 3D displacement is reconstructed using a triangulation technique. This system has several advantages over traditional cable-based stereo vision systems for tall structures: 1) The distance between the two camera nodes is not limited, allowing flexible deployment to optimise measurement accuracy for structures of varying heights. 2) It avoids transmitting large volumes of image data between the cameras and the 3D image-processing computer, reducing the need for high network bandwidth. 3) It simplifies the stereo calibration process by eliminating the need for a checkerboard, which is often impractical to be positioned in the field of view of the cameras in tall structure applications. An outdoor test was conducted to validate the system, with the cameras placed about 200 m from the target and 100 m apart. The results showed a measurement accuracy of approximately 1 mm within the horizontal plane.

KEY WORDS: Computer Vision, Structural Health Monitoring, Stereo Vision, Wireless Sensor, 3D Displacement Measurement.

## 1 INTRODUCTION

Vision-based systems provide a contactless alternative to traditional sensors. They enable the collection of structural displacement data for structural health monitoring (SHM). As a result, these systems have become popular in the SHM field. The imaging process projects objects in 3D space onto a 2D image plane [1]. Therefore, a single camera can only measure 2D displacement unless special calibration patterns are used.

To measure displacement in 3D, two cameras are combined into a stereo vision system. The depth information of objects is obtained through triangulation. Usually, both cameras are connected to a common controller computer via Ethernet cables. Before measurement, stereo calibration is performed to determine the essential matrix between the images from the two cameras. Next, the controller triggers and synchronizes the two cameras simultaneously to capture images. These images are then transmitted to the controller for processing, including stereo rectification, stereo correspondence, and 3D point reconstruction. Stereo vision systems have been successfully applied to measure the 3D displacement of structures such as wind turbines [2] and bridges [3].

However, measurement accuracy in depth decreases when the measurement distance increases if the distance between the two cameras (the baseline) remains constant. For tall structures, the distance between cameras and structures can be significant. To maintain measurement accuracy in depth, a baseline of potentially hundreds of meters is required. In such cases, conventional cable-based stereo vision systems become impractical.

To address this limitation, cable-free stereo vision systems have been developed. Yang et. al [4] introduced a mobile stereo vision system with an adjustable baseline distance. Images from the cameras are transmitted wirelessly to a controller computer. Sumetheeprasit et .al [5] developed a stereo vision system using two drone-mounted cameras with a variable baseline. However, existing cable-free stereo vision systems mainly target short-term, offline measurements. Long-term online monitoring poses additional challenges, such as ensuring long-term synchronized image capture, managing the large bandwidth needed for wireless image transmission, and performing real-time stereo image processing. These challenges complicate the long-term online monitoring of tall structures.

This paper presents a novel wireless stereo vision system designed for long-term online monitoring of 3D displacement in tall structures. Figure 1 illustrates the overall workflow of this system:

1) The left and right cameras are strategically positioned to target a common point on the structure.

2) Stereo calibration is performed by measuring the relative geometrical relationship (the essential matrix) between the two cameras using GPS-RTK. A world coordinate system is established, with the origin located at the center of the left camera. The Z-axis extends from the center of the left camera toward the center of the right camera. The Y-axis points vertically upward, and the X-axis is defined according to the right-hand rule. This approach simplifies coordinate system transformation during 3D reconstruction.

3) Both cameras are simultaneously triggered by 10 Hz pulse-per-second (PPS)-based signals to capture synchronized images.

4) Reference images from both cameras are selected, and feature points are detected using the deep learning algorithm SuperPoint [6]. These detected points are matched using the LightGlue algorithm [7]. Bounding boxes around these matched point pairs serve as candidates for selecting a tracking target using the zero-normalized cross-correlation (ZNCC) algorithm.

5) The cameras capture images every half hour or one hour to document illumination changes throughout the day. The ZNCC scores of candidate targets across these images are calculated, and the bounding box with the highest ZNCC score is selected for long-term tracking.

6) Measurement begins, with each camera independently processing its captured images to determine the 2D pixel position of the selected target.

DOI: 10.3217/978-3-99161-057-1-051

7) The extracted 2D data from each camera is automatically uploaded to Amazon Web Services (AWS) S3 for storage.

8) AWS Lambda computes the 3D displacement using the uploaded 2D data from both cameras.

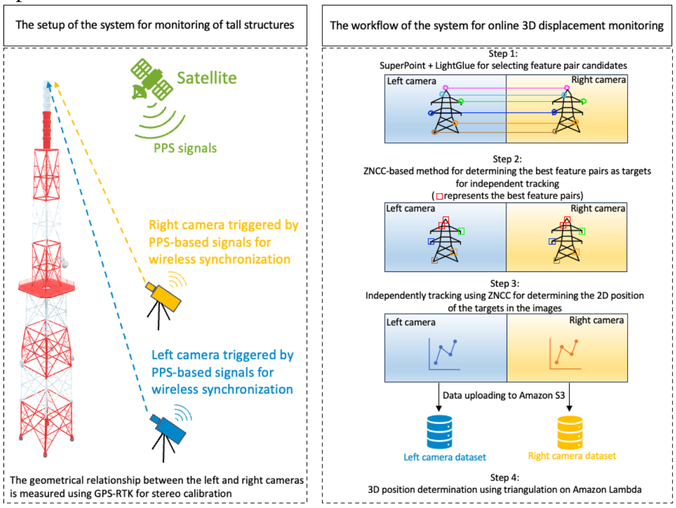



Figure 1. The workflow of the wireless stereo vision system.

## 2 WIRELESS STEREO VISION SYSTEM

### 2.1 Hardware

The wireless stereo vision system consists of left and right camera nodes, each with identical hardware components. Each node includes a SONY IMX296 camera with a lens, two Raspberry Pi 4B computers, a GNSS board, a GNSS antenna, and a 4G module. The camera can be triggered by external signals. To wirelessly trigger both cameras simultaneously, PPS signals are used. Once the first Raspberry Pi receives the PPS signal from satellites, it generates a 10 Hz trigger signal to control the camera's image capture. The PPS signal from the satellites occurs precisely at the start of every second, providing nanosecond-level accuracy. To ensure durability for long-term field operation, the hardware components of each camera node are enclosed in a CCTV camera housing.

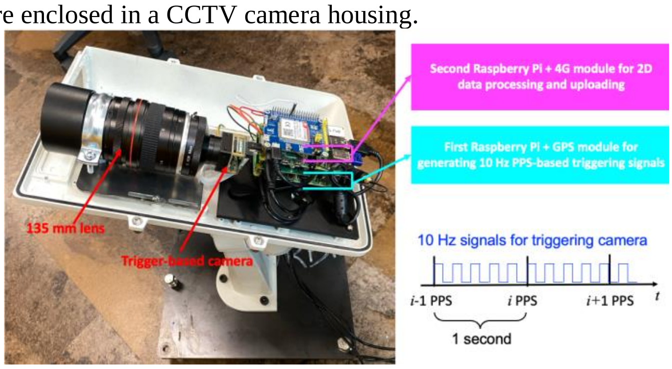



Figure 2. The hardware of the left/right camera.

### 2.2 software

The software workflow for online 3D displacement monitoring is shown in Figure 1 and described in the Introduction. This section explains how 3D displacement is calculated from the 2D data captured by the two cameras.

In the proposed system, unlike conventional stereo vision systems, the left and right cameras are not permanently fixed during manufacturing. Instead, their relative positions can be adjusted according to different scenarios. Therefore, stereo calibration must be performed before each measurement. Stereo calibration determines the essential matrix, which describes the geometric relationship between the image planes of the two cameras. This essential matrix combines the intrinsic parameters of the two cameras and a fundamental matrix.

Typically, the two cameras are placed far apart, with a large overlapping field of view (FOV), aimed at the top of the monitored structures. In this setup, using a calibration pattern to directly obtain the essential matrix is challenging. As a result, the intrinsic camera matrices and the fundamental matrix are calibrated separately.

The intrinsic matrices of the two cameras, K_*left* and K_*right*, can be obtained using Zhang's calibration method [8]. The fundamental matrix is calculated using GPS-RTK, as shown in Figure 3. The calibration process for the fundamental matrix follows these steps:

1) The coordinates of the left and right cameras, as well as the structure's ground level, are measured using GPS-RTK. A world coordinate system is defined with the origin at the center of the left camera. The camera centers are adjusted to point toward the target, and the roll angles are set to zero. The yaw angles of both cameras are then calculated. The left camera is located at (0, 0, 0), and the right camera at (0, V, S).

2) The target's coordinates in the world coordinate system are determined based on structural drawings. The pitch angles of the cameras are then calculated. The rotation matrix from the left camera's coordinate system to the world coordinate system is: R_*left* = R(Yaw_*left*) R(Pitch_*left*) R(Roll_*left*), where

$$R(Yaw\_left) = \begin{bmatrix} \cos(Yaw\_left) & -\sin(Yaw\_left) & 0 \\ \sin(Yaw\_left) & \cos(Yaw\_left) & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$R(Pitch\_left) = \begin{bmatrix} \cos(Pitch\_left) & 0 & \sin(Pitch\_left) \\ 0 & 1 & 0 \\ -\sin(Pitch\_left) & 0 & \cos(Pitch\_left) \end{bmatrix}$$

$$R(Roll\_left) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(Roll\_left) & -\sin(Roll\_left) \\ 0 & \sin(Roll\_left) & \cos(Roll\_left) \end{bmatrix}.$$

The translation vector from the left camera to the world coordinate system is T_*left* = (0, 0, 0). Similarly, the rotation matrix and translation vector for the right camera are: R_*right* = R(Yaw_*right*) R(Pitch_*right*) R(Roll_*right*) and T_*right* = (0, V, S).

3) The fundamental matrix F between the left and right cameras is calculated by $F = (R\_right\ T\_right)^{-1} (R\_left\ T\_left)$. Then the essential matrix E is obtained as $E = (K\_right)^{T} F\ K\_left$.

DOI: 10.3217/978-3-99161-057-1-051

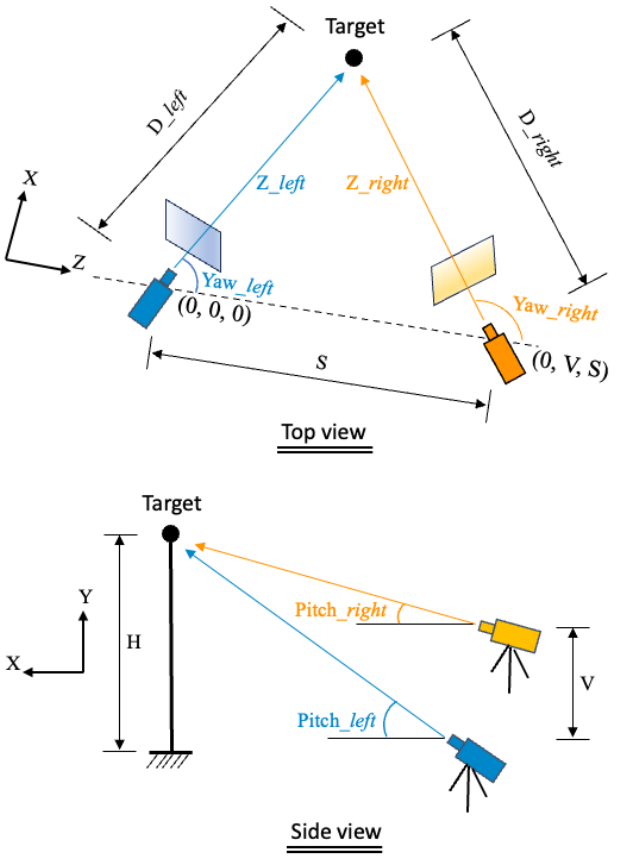
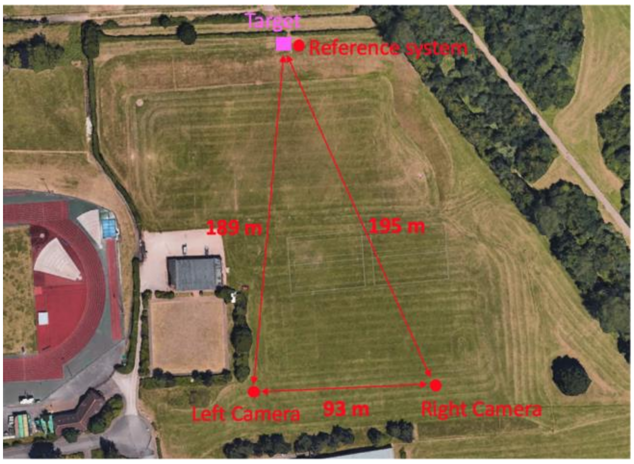



Figure 3. The determination of the foundation matrix.

## 3 VALIDATION TEST

To evaluate the measurement accuracy of the system, an outdoor experiment was conducted. The experimental setup is shown in Figure 4. A vertical cantilever, 2 meters in height, was used as the measurement target. The left and right camera nodes were placed at long distances from the cantilever—189 meters and 195 meters away, respectively. The baseline distance between the two cameras was 93 meters.




(a)


(b)

(c)

Figure 4. The setup of the validation experiment.

The stereo vision system measured the 3D displacement of an artificial target (concentric circles) placed at the top of the cantilever. For reference, the horizontal displacement was also measured using a separate 2D vision-based system. A chessboard target was used for this reference measurement.

The results from the stereo vision system were compared with those from the 2D reference system, as shown in Figure 5. The average root mean square error (RMSE) between the two measurements was 1.02 mm.

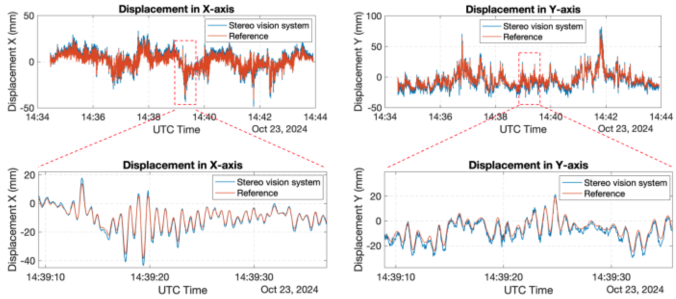



Figure 5. The measurement results of the validation experiment.

## REFERENCES

[1] M.M. Wang, W.K. Ao, J. Bownjohn, F.Y. Xu, A novel gradient-based matching via voting technique for vision-based structural displacement measurement, Mech Syst Signal Pr, 171 (2022).

[2] W. Feng, Q. Li, W. Du, D. Zhang, Remote 3D displacement sensing for large structures with stereo digital image correlation, Remote Sensing, 15 (2023) 1591.

[3] X.-H. Long, W.-W. Zhan, X.-L. Gui, Research on displacement monitoring method of heavy load bridge structure based on stereo vision, 2020 IEEE

DOI: 10.3217/978-3-99161-057-1-051

International Conference on Industrial Application of Artificial Intelligence (IAAI), IEEE, 2020, pp. 432-437.
[4] Y. Wang, X. Wang, A mobile stereo vision system with variable baseline distance for three-dimensional coordinate measurement in large FOV, Measurement, 175 (2021) 109086.
[5] B. Sumetheeprasit, R. Rosales Martinez, H. Paul, R. Ladig, K. Shimonomura, Variable baseline and flexible configuration stereo vision using two aerial robots, Sensors, 23 (2023) 1134.
[6] D. DeTone, T. Malisiewicz, A. Rabinovich, Superpoint: Self-supervised interest point detection and description, Proceedings of the IEEE conference on computer vision and pattern recognition workshops, 2018, pp. 224-236.
[7] P. Lindenberger, P.-E. Sarlin, M. Pollefeys, Lightglue: Local feature matching at light speed, Proceedings of the IEEE/CVF international conference on computer vision, 2023, pp. 17627-17638.
[8] Z. Zhang, A flexible new technique for camera calibration, IEEE Transactions on pattern analysis and machine intelligence, 22 (2000) 1330-1334.