

Article

Vision-Based Building Seismic Displacement Measurement by Stratification of Projective Rectification Using Lines

Jia Guo ^{1,*} , Yang Xiang ^{2,†} , Kohei Fujita ² and Izuru Takewaki ² ¹ International Research Institute of Disaster Science (IRIDeS), Tohoku University, Sendai 980-8572, Japan² Department of Architecture and Architectural Engineering, Kyoto University, Kyoto 615-8540, Japan; xiang.ya@m.titech.ac.jp (Y.X.); fm.fujita@archi.kyoto-u.ac.jp (K.F.); takewaki@archi.kyoto-u.ac.jp (I.T.)

* Correspondence: guo@irides.tohoku.ac.jp

† Current address: Institute of Innovative Research (IIR), Tokyo Institute of Technology, Yokohama 226-8503, Japan.

Received: 18 September 2020; Accepted: 5 October 2020; Published: 12 October 2020



Abstract: We propose a new flexible technique for accurate vision-based seismic displacement measurement of building structures via a single non-stationary camera with any perspective view. No a priori information about the camera's parameters or only partial knowledge of the internal camera parameters is required, and geometric constraints in the world coordinate system are employed for projective rectification in this research. Whereas most projective rectifications are conducted by specifying the positions of four or more fixed reference points, our method adopts a stratified approach to partially determine the projective transformation from line-based geometric relationships on the world plane. Since line features are natural and plentiful in a man-made architectural building environment, robust estimation techniques for automatic projective/affine distortion removal can be applied in a more practical way. Both simulations and real-recorded data were used to verify the effectiveness and robustness of the proposed method. We hope that the proposed method could advance the consumer-grade camera system for vision-based structural measurement one more step, from laboratory environments to real-world structural health monitoring systems.

Keywords: seismic displacement measurement; vision-based system; projective rectification; vanishing line estimation

1. Introduction

Observing the vibration characteristics of building structures in a metropolitan area subjected to extreme loading (e.g., earthquake) with state-of-the-art sensing technologies has become a valuable method for structural safety assessments and design validations. Generally, the behavior of building structures could be evaluated in terms of two types: the acceleration response corresponding to the short period components of the external load and the displacement response corresponding to the long period components. The monitoring of the short-term as well as of long-term responses of building structures under severe loading conditions and verifying their spectral characteristics are both of importance in many cases. For instance, high-rise buildings far from the epicenter of the Great East Japan earthquake were reported to have been shaken intensively by the ground motion of short to long period components [1,2]. In practice, acceleration measurements are most convenient for use in structural dynamics applications. Nevertheless, monitoring the displacement responses of structures via the double integration of accelerations is always problematic due to the nature of signal processing: the integration constants cannot be determined, and thus only part of the response can be fully

recovered. Besides, the numerical integration itself can also be inaccurate, based on the theory of error propagation [3]. With the recent development of sensing and robotic technologies, new generation displacement sensing techniques, including Global Positioning System (GPS), Speckle Photography, and laser Doppler vibrometer, provide significant advantages to supplement accelerometers for structural response monitoring. However, utilization of these techniques has to overcome many practical limitations such as the issue of obstructions, the requirement of high set costs, the need of stable platform, complex data interpretation, and hardly fully assessed accuracy [4–6].

Recently, with the aid of the digital image processing techniques and modern computer and information science, the field of vision-based displacement measurement techniques has attracted intensive attention in structural health monitoring, for its advantages of ease of operation and non-contact flexibility. A vision-based measurement system typically consists of one video camera or multiple cameras, and a computer. While most previous vision-based applications in structural response monitoring end up with producing satisfied displacement tracking results using stationary camera [7–9], which is not easy to ensure for field measurements during earthquakes or typhoons, limited research examined the effects of the camera motion on the measurement accuracy.

A first general approach to consider for freely moving camera is the computation of an optimal visual reconstruction of camera motion and 3D scene structure by extracting background point-based features within the un-deformed/rigid region of the view. A typical application of this method in structural vibration-based monitoring can be found in [10]. Challenging tasks during the reconstruction process involve the issue of matching ambiguities for repeated structures in the scene (e.g., regular building facade), and achieving sub-pixel measurement accuracy, since small deviations in the camera motion estimation might result in large measurement errors.

An alternative simplified approach to address this challenge was to explore partial or complete 3D metric information directly from the analysis of 2D geometric properties in the world system. When a projection of 2D world coordinates to 2D image coordinates is obtained, the effect of the camera motion can be directly removed by the projective rectification. This approach is always precise enough, easy to conduct and thus widely applied in laboratory experiments and field tests [11–13]. In order to improve the accuracy of image rectification, enough high-contrast natural regions of interest (ROIs) from parallel background world planes are required to compute the projective transformation for each image. However, recent research shows that these ROIs are always deficient in real-world applications [14–16].

Meanwhile, for a wide variety of man-made architectural building environments, *line segments* have been proven useful to achieve rich and reliable geometric information of the 3D structure in images, due to their abundance on man-made objects. Examples and applications of line-based vision task include facade detection [17], camera calibration [18], camera pose estimation [19], stereo SLAM [20], etc. State-of-the-art technologies with high-precision as well as full automatization in line detection [21], line matching [22], vanishing point detection [23], etc., provide solid basic steps for the completion of these tasks.

In this paper, we propose a novel vision-based seismic displacement measurement method designed for building structures. We address the problem of projective transformation estimation by extracting line-based, rather than point-based, geometric information from the architectural building environments. Since line segments are always abundant in building structures, robustness estimation techniques for automatic projective/affine distortion removal can be applied in a more practical way. The main advantages of our proposed algorithm include:

- This study addresses the problem of extracting structural dynamic displacement information from a *single, uncalibrated* camera;
- There is no need for stationary cameras and cameras with any perspective view are allowed to be used during the measurement process;
- Line segments are natural and plentiful in man-made architectural buildings, which makes the proposed algorithm applicable in real-world applications;

- The proposed algorithm is especially useful for *automatic* perspective distortion removal and image rectification from video sequence.

Unlike the existing image rectification techniques [24,25], we decompose the projective transformation into a chain of transformations, i.e., the pure projective transformation, the affine transformation and the similarity transformation. This process is defined as “stratification of projective rectification”. With this idea the problem can be stratified into different steps and in each step, only two or three rectification parameters have to be solved from nonlinear equations. We employ parallelism and orthogonality relationships of lines that are common and plentiful in architectural scenes, to obtain projective and then affine rectification parameters. With these parameters determined, images could be rectified up to a similarity and the relative positions, displacements and deformations of the building can be finally computed by a global scale factor.

The remainder of this paper is structured as follows: Section 2 lays the foundation for the representations and terminology that will be used in the subsequent parts of this paper and introduces the basic idea of line-based projective rectification. Section 3 provides the experimental results with both synthetic images (Section 3.1) and real videos (Section 3.2). The final conclusions are drawn in Section 4.

2. Methodology

Notations and mathematical operations in this study are defined as follows. Scalars are denoted as (x, a) while vectors and matrices are denoted as $(\mathbf{x}, \mathbf{M})$. $\otimes$ is the Kronecker product. $\text{vec}(\cdot)$ is used to represent the linear transformation which converts the matrix into a column vector. The cross product between vectors $\mathbf{a} = (a_1, a_2, a_3)$ and $\mathbf{b} = (b_1, b_2, b_3)$ can be written as matrix multiplications $\mathbf{a} \times \mathbf{b} = [\mathbf{a}]_{\times} \mathbf{b}$, where $[\mathbf{a}]_{\times}$ is defined as

$$[\mathbf{a}]_{\times} = \begin{bmatrix} 0 & -a_3 & a_2 \\ a_3 & 0 & -a_1 \\ -a_2 & a_1 & 0 \end{bmatrix} \quad (1)$$

We also introduce the basic idea of projective transformation and some important notations herein. One 3D coordinate system and two 2D coordinate systems are defined in this study: a 3D world coordinate system, where points are denoted by homogeneous coordinates with upper case letters $\mathbf{X} = (X, Y, Z, 1)^T$, a 2D world planar system, where we assume the plane is on $Z = 0$ of the world coordinate system without loss of generality, and an image plane coordinate system, where points are denoted by lower case letters $\mathbf{x} = (x, y, 1)^T$. Correspondence entities from different coordinate systems have the same subscript, e.g., $\mathbf{L}_{\lambda} \Leftrightarrow \mathbf{l}_{\lambda}$.

A camera is modeled by the usual pinhole. The transition from the world coordinate system to the image plane coordinate system for each camera is realized through a translation followed by a rotation. Let a 3×1 translation vector $\mathbf{C} = (C_x, C_y, C_z)^T$ represent the position of the camera center in the world coordinate system and a 3×3 rotation matrix $\mathbf{R}$ describe the orientation of the camera by means of three consecutive rotations along the three axes Z, Y, X by respective Euler angles Γ, B, A . The parameters from $\mathbf{C}$ and $\mathbf{R}$ are called camera extrinsic parameters. As regards the intrinsic parameters of the pinhole camera, intrinsic matrix $\mathbf{K}$ is given by

$$\mathbf{K} = \begin{bmatrix} f_x & s & p_x \\ 0 & f_y & p_y \\ 0 & 0 & 1 \end{bmatrix} \quad (2)$$

where $[p_x, p_y]$ is the coordinate of the principle point and $[f_x, f_y]$ represents the focal length of the camera in terms of pixel dimensions. s is the skew parameter. Once the camera parameters are

determined, the camera projection matrix, defined as $\mathbf{P}_c = \mathbf{K}\mathbf{R}[\mathbf{I}|\mathbf{C}] = \mathbf{K}[\mathbf{R}|\mathbf{t}]$, where $\mathbf{t} = \mathbf{RC}$, can be computed. The relationship between a 3D point $\mathbf{X}$ and its image projection $\mathbf{x}$ is consequently given by

$$\mathbf{x} = \mathbf{P}_c \mathbf{X} \quad (3)$$

Let us denote the i th column of the rotation matrix $\mathbf{R}$ by $\mathbf{r}_i$. The 3D world point $\mathbf{X}$ on the world plane $Z = 0$ has the form $\hat{\mathbf{X}} = (X, Y, 0, 1)$ and is defined by just two coordinates X and Y . It is projected into image plane system via $\mathbf{P}_c$, yielding the 2D point $\mathbf{x}$

$$\mathbf{x} = \mathbf{K} \begin{bmatrix} \mathbf{r}_1 & \mathbf{r}_2 & \mathbf{r}_3 & \mathbf{t} \end{bmatrix} \begin{bmatrix} X \\ Y \\ 0 \\ 1 \end{bmatrix} = \mathbf{K} \begin{bmatrix} \mathbf{r}_1 & \mathbf{r}_2 & \mathbf{t} \end{bmatrix} \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} = \mathbf{H}' \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} \quad (4)$$

where the non-singular homogeneous matrix $\mathbf{H}' \in \mathcal{R}^{3 \times 3}$ is the projective transformation between $\mathbf{x}$ and $\hat{\mathbf{X}}$. It can be seen that there is a bijection between the image plane and the world plane. That is to say, point $\mathbf{x}$ on the image plane is paired with exactly one point $\hat{\mathbf{X}}$ on the world plane by $(\mathbf{H}')^{-1} = \mathbf{H}$. For the purpose of clarity, we slightly abuse the notation $\mathbf{X}$ to denote $\hat{\mathbf{X}}$, i.e., $\mathbf{X} = (X, Y, 1)^T$. Then, the relation between $\mathbf{X}$ and $\mathbf{x}$ becomes

$$\mathbf{X} = \mathbf{H}\mathbf{x} \quad (5)$$

Once $\mathbf{H}$ is obtained, any image point can be mapped into the corresponding location on the world plane $Z = 0$. That is to say, the (X, Y) coordinate of any tracked point in the 3D world system, whether on the plane $Z = 0$ or not, is able to be reconstructed. Thus, the only remaining problem is that of estimating the projective transformation $\mathbf{H}$. Normally, $\mathbf{H}$ is computed directly from a set of at least four corresponding points. In this research, however, line-based geometric relationships on the world plane, such as parallelism and orthogonality, are employed to compute the projective and affine components of $\mathbf{H}$.

The point $\mathbf{x} = (x, y, 1)^T$ lies on the line $\mathbf{l} = (a, b, c)^T$ if and only if $\mathbf{x}^T \mathbf{l} = ax + by + c = 0$. Under the point transformation $\mathbf{X} = \mathbf{H}\mathbf{x}$, a line transformation can be written in the form

$$\mathbf{L} = \mathbf{H}^{-T} \mathbf{l}. \quad (6)$$

On the image plane system, a vanishing point is a point where mutually parallel lines in the world system appear to converge. For different sets of lines parallel to a world plane, their respective vanishing points may lie on a same line, called a *vanishing line*, which is expressed as $\mathbf{l}_\infty$ in this research. Under a projective transformation, vanishing points are finite points, and consequently $\mathbf{l}_\infty$ is mapped to a finite line. Understanding the meaning of vanishing line is important because, in what follows, it is shown that the projective distortion may be removed once $\mathbf{l}_\infty$ is specified.

As shown in Figure 1, our proposed technique works in five steps: (1) image segmentation, (2) line detection and segment clustering, (3) vanishing line estimation, (4) stratification of projective rectification, and (5) displacement measurement. In the rest of this section, we provide technical details for steps (1)–(4). Having obtained the projective transformation $\mathbf{H}$, any image-tracked point can be projected to the world plane and the displacement of this point in plane is then obtained by calculating the Euclidean distance between this point and the stationary point with a global scale factor. Since this process in step (5) is rather clear, it will not be discussed in detail in this section.

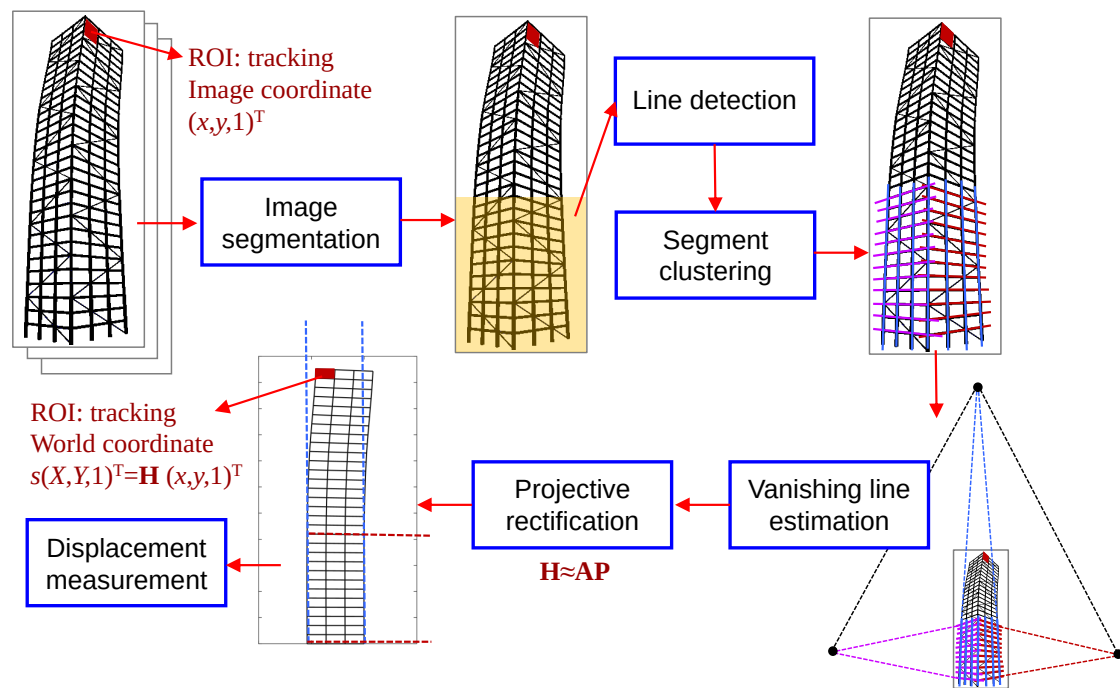


Figure 1. Overview of the proposed technique.

2.1. Image Segmentation

Image segmentation is always an essential component in many vision-based systems. It involves partitioning images into multiple objects [17]. For the case of structural displacement measurement in this study, image segmentation is adopted first to identify the target building facade from surrounding urban areas. Next, the target building facade is divided into two parts: the (nearly) rigid region (part I) and the deformed region (part II) of the building, as shown in Figure 2.

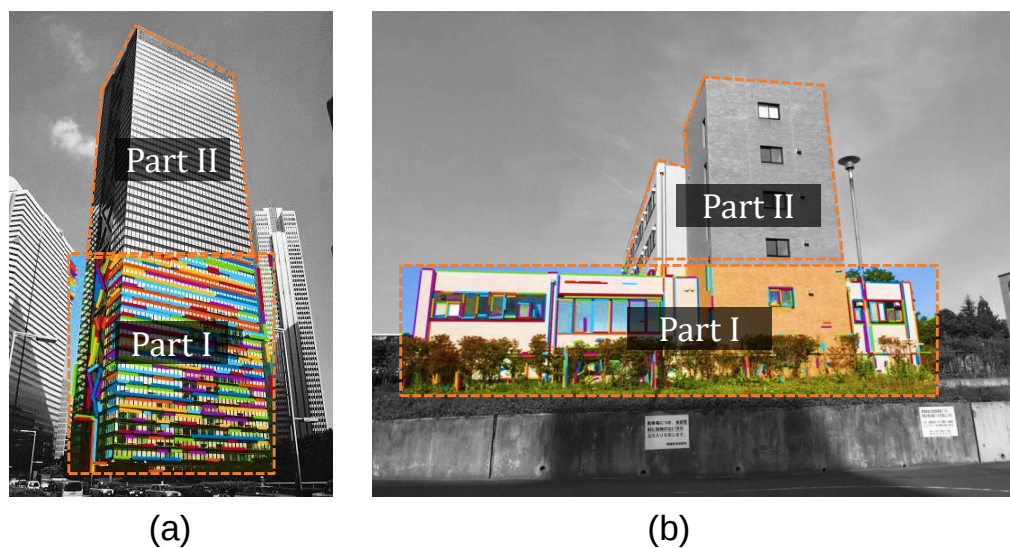


Figure 2. Examples of image segmentation (Line detection by [21]): (a) target building itself as part I; (b) structures from surrounding environments as part I.

Numerous image segmentation algorithms have been proposed in the literature for building facade segmentation. Different techniques should be chosen for different field situations: from the earliest thresholding [26], k-means clustering [27] methods to the most popular deep learning-based methods [28]. It is noteworthy that automatic building facade detection methods [29–31] have been well developed for geometric 3D reconstruction, and thus are more suitable for the application in the proposed vision-based measurement system.

After successful identification of the target building from the surrounding structures, line detection and projective rectification steps are restricted to part I of the segmentation. The rigid region part I is only an approximation and may vary in building types, building heights and surrounding environments. Any structures which have negligible deformations compared to the target building can be included, as shown in Figure 2b. Section 3.1 also proved that even the lower part of the target building itself (as in Figure 2a) could be set as the rigid region.

2.2. Line Detection and Segment Clustering

Building facades generally exhibit numerous lines that are either parallel or orthogonal to the gravity direction. Many methods have been proposed for automatic line segment detection since the 1980s, such as Canny operator [32] and Hough transform [33]. These methods are generally slow and usually combine non-contiguous line segments together, producing a lot of false detections. The recently proposed fast, parameterless line detectors, e.g., EDLines [34] and LSD [35], have shown good detection results for most types of image with enough accuracy and robustness and have been widely employed for line-based vision tasks.

Given a set of lines extracted from the building facade, the existing methods mainly rely on clustering these lines into sets of world parallel lines that converge to the same vanishing points. Commonly, the RANSAC estimation algorithm to search for concurrent lines in the image is used for segment clustering. In the RANSAC algorithm, the fitting model consists of a point and the sample is obtained by choosing randomly pairs of line segments. The intersection of each pair is computed so as to get putative vanishing points, and then the support for this vanishing point is found. The final concurrent lines result from the model with sufficient supports. This procedure was described in detail in [36]. Other segment clustering methods include identifying different groups of parallel lines in the PCLines dual spaces [37] or relying on the Helmholtz principle [38].

Two groups of line segments are expected to be obtained for each of the target building facade plane by the above segment clustering strategies: the vertical lines $\mathbf{l}_z$ and the horizontal lines $\mathbf{l}_h$.

2.3. Vanishing Line Estimation

Normally, vanishing points at different directions $\mathbf{v}_h, \mathbf{v}_z$, which correspond to the horizontal lines $\mathbf{l}_h$ and the vertical lines $\mathbf{l}_z$ respectively, can be found once the segment clustering has been determined. Two vanishing points determine a vanishing line $\mathbf{l}_\infty$, i.e., $\mathbf{l}_\infty = \mathbf{v}_h \times \mathbf{v}_z$, which is of key importance to extract geometric constraint information and remove projective distortion from the current image, as shown in Figure 3a. However, challenges in vanishing point detection arise from the inherent measurement error. The existing pixel noise, image distortion and discretization error might greatly affect the location of the vanishing point greatly, especially when camera parameters and motions are unavailable, or the vertical vanishing point is found to be infinite, when the problem becomes even harder.

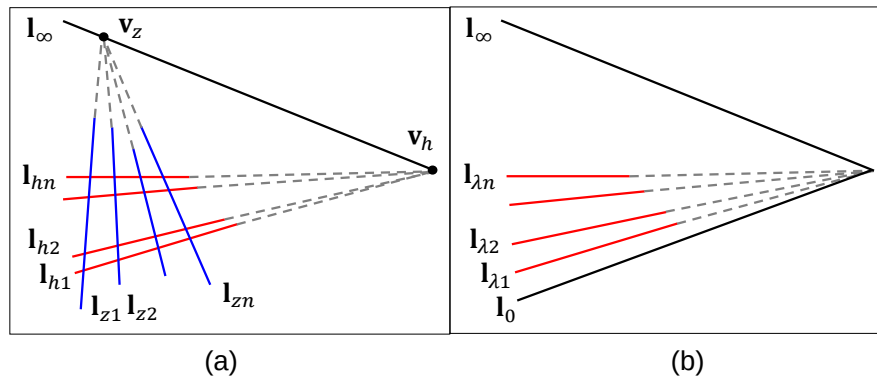


Figure 3. Vanishing line estimation: (a) method using vanishing points, (b) method using equally spaced parallel lines.

If part of the horizontal lines corresponding to $\mathbf{l}_h$ are equally spaced in the world coordinate system, which is easily satisfied for building facades with regular textures, an alternative vanishing line estimation method can be applied without the need of estimates of vanishing points. Consider a group of parallel lines $\mathbf{L}_\lambda$ on the world plane $Z = 0$ with the following uniform expression:

$$\mathbf{L}_\lambda : aX + bY + \lambda = 0 \quad (7)$$

This group of parallel lines may naturally be represented by the common normal vector $(a, b)^T$ and different lines have different values of λ . With this preparation and further considering Equation (6), one could obtain the corresponding lines in the image plane as

$$\mathbf{l}_\lambda = \mathbf{H}^T \mathbf{L}_\lambda = \mathbf{H}^T \begin{pmatrix} 0 & a \\ 0 & b \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \lambda \\ 1 \end{pmatrix} = \mathbf{V} \begin{pmatrix} \lambda \\ 1 \end{pmatrix} \quad (8)$$

where $\mathbf{V}$ is a non-singular 3×2 matrix and determined to scale. It has been proved that the first column of $\mathbf{V}$ stands for the vanishing line $\mathbf{l}_\infty$ and the second column of $\mathbf{V}$ represents the line $\mathbf{l}_0$ [36], as shown in Figure 3b. Consequently, once the matrix $\mathbf{V}$ is determined, the vanishing line can be obtained. The process of estimating the matrix $\mathbf{V}$ through line correspondences ($\lambda_i \Leftrightarrow \mathbf{l}_{\lambda_i}, i = 1, \dots, n$) is similar to that of estimating homography $\mathbf{H}$ through point correspondences: we begin with the basic Direct Linear Transformation (DLT) algorithm to compute the initial value of $\mathbf{V}$ and then turn to the iterative minimization method to find the optimal estimate of $\mathbf{V}$.

DLT algorithm. Now, swapping both sides of Equation (8) and multiplying them by $[\mathbf{l}_\lambda]_\times$ yields

$$[\mathbf{l}_\lambda]_\times \mathbf{V} \begin{pmatrix} \lambda \\ 1 \end{pmatrix} = \mathbf{0} \quad (9)$$

Using the Lemma in [39], one has

$$((\lambda, 1) \otimes [\mathbf{l}_\lambda]_\times) \cdot \text{vec}(\mathbf{V}) = \mathbf{0} \quad (10)$$

If we set $\mathbf{M}_i = (\lambda_i, 1)_i \otimes [\mathbf{l}_{\lambda_i}]_\times$ and $\mathbf{a} = \text{vec}(\mathbf{V})$, Equation (10) can be re-written as

$$\mathbf{M}_i \mathbf{a} = \mathbf{0} \quad (11)$$

Each line correspondence gives rise to two independent equations of $\mathbf{a}$. Given a set of three such line correspondences, we obtain a set of equations $\mathbf{M} \mathbf{a} = \mathbf{0}$, where $\mathbf{M}$ is the matrix of equation coefficients built from the matrix rows $\mathbf{M}_i$. Note that the $\mathbf{M}$ has rank 5, and thus has a 1-dimensional

null-space which provides a solution for $\mathbf{a}$. Generally, $\mathbf{a}$ is obtained by the singular value decomposition (SVD) of $\mathbf{M}$, i.e., the unit singular vector corresponding to the smallest singular value of $\mathbf{M}$ is the solution $\mathbf{a}$.

Iterative minimization. With the initial value of $\mathbf{V}$ obtained from the DLT algorithm, the next step is to get the best estimate of $\mathbf{V}$ by minimizing an appropriate geometric cost function with the use of iterative techniques. Herein, the cost function is selected as

$$\sum_{i=1}^n \sum_{j=1}^m d(\mathbf{x}_{ij}, \mathbf{l}_{\lambda i}(\mathbf{V}))^2 \quad (12)$$

where $d(\mathbf{x}, \mathbf{l})$ represents the perpendicular distance from a point $\mathbf{x}$ to the line $\mathbf{l}$. $\mathbf{l}_{\lambda i}(\mathbf{V}) = \mathbf{V}(\lambda_i, 1)^T$ and $\mathbf{x}_{ij}$ is the j th ($j = 1, \dots, m$) point on the i th line on the image plane. In this research, the cost function in Equation (12) is minimized using the Levenberg–Marquardt algorithm.

Grouping the equally spaced lines $\mathbf{l}_{\lambda}$ from $\mathbf{l}_h$ can also be addressed by the RANSAC algorithm. For a complete exposition of this procedure, the reader is referred to [36].

2.4. Stratification of Projective Rectification

The goal of projective rectification is to remove the projective/affine distortion in the original image plane (π_1) to the extent that similarity properties (angles, ratios of lengths) could be measured on the transformed plane (π_3). In most of the previous research, this was directly completed by specifying the position of at least four reference points. In this research, however, metric structure recovery is stratified so that the pure projective distortion is removed first and then the affine distortion is corrected, by using line-based constraints (i.e., parallelism and orthogonality). The idea of stratified projective rectification is illustrated in Figure 4.

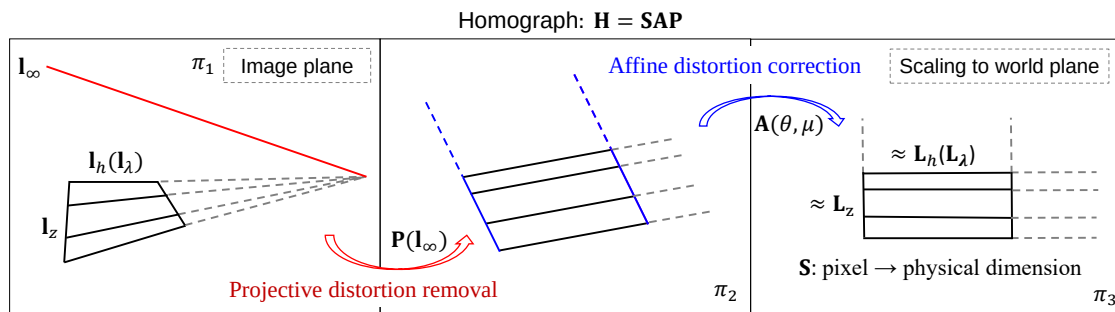


Figure 4. Image rectification: from projective distortion removal to affine distortion correction.

For the group of invertible $n \times n$ matrices with real elements, the general linear group on n dimensions can be expressed as $GL(n)$. The homography is a quotient group of $GL(3)$, giving $PL(3)$ (e.g., $\det(\mathbf{H}) = 1$). According to [40], subgroups of $PL(3)$ include the affine group and the similarity group. As a result, the projective transformation matrix $\mathbf{H}$ could be uniquely decomposed into three matrices [41]

$$\mathbf{H} = \mathbf{SAP} \quad (13)$$

where $\mathbf{P}$ is a ‘pure projective’ transformation, $\mathbf{A}$ is an affine transformation and $\mathbf{S}$ is a similarity transformation, respectively.

The ‘pure projective’ transformation $\mathbf{P}$ can be obtained directly by the vanishing line $\mathbf{l}_{\infty} = (l_1, l_2, l_3)^T$ of the image plane as

$$\mathbf{P} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ l_1 & l_2 & l_3 \end{pmatrix} \quad (14)$$

Under a projective transformation, $\mathbf{l}_\infty$ is mapped to a finite line while under an affine transformation, $\mathbf{l}_\infty$ is not mapped to a finite line but remains at infinity. In other words, if the imaged line at infinity is the line $\mathbf{l}_\infty = (l_1, l_2, l_3)^T$, then provided $l_3 \neq 0$, the above ‘pure projective’ $\mathbf{P}$ maps the line $\mathbf{l}_\infty$ on π_1 back to the line $(0, 0, 1)^T$ on a new plane π_2 , where no projective distortion exists. This is directly evident from Equation (6) that $\mathbf{P}^{-T}(l_1, l_2, l_3)^T = (0, 0, 1)^T$.

The affine transformation in Equation (13) is represented by an upper-triangular matrix

$$\mathbf{A} = \begin{pmatrix} \frac{1}{\beta} & -\frac{\alpha}{\beta} & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (15)$$

where parameters α and β represent the circular points of the plane, which are invariant under similarity transformations. The circular point is a pair of complex conjugate points on the vanishing line. Circular points are transformed from similarity coordinates $(1, \pm i, 0)^T$ to affine coordinates $(\alpha \mp \beta i, 1, 0)^T$ by the affine transformation $\mathbf{A}$.

Finally, the similarity transformation $\mathbf{S}$ is expressed by

$$\mathbf{S} = \begin{pmatrix} s\mathbf{R} & \mathbf{t} \\ \mathbf{0}^T & 1 \end{pmatrix} \quad (16)$$

where $\mathbf{R}$ is a 2×2 orthogonal rotation matrix, s an scaling factor, $\mathbf{t}$ a translation 2-vector and $\mathbf{0}$ a null 2-vector.

We ignore the metric part $\mathbf{S}$ of $\mathbf{H}$ in the solving process of projective rectification and only recover the non-metric part $\mathbf{N} = \mathbf{A}\mathbf{P}$. Under a stratified rectification scheme, the two components, $\mathbf{P}$ and $\mathbf{A}$ of $\mathbf{N}$, are recovered step by step, as follows.

2.4.1. Projective Distortion Removal

Based on Equation (14), the pure projective transformation $\mathbf{P}$ is only determined by the vanishing line $\mathbf{l}_\infty$. As mentioned in Section 2.3, two methods may be used to determine $\mathbf{l}_\infty$: the method using vanishing points $\mathbf{v}_h, \mathbf{v}_z$ and the method using equally spaced parallel lines $\mathbf{l}_\lambda$. We summarize details for each of the methods to obtain the transformation $\mathbf{P}$ in Table 1 (Part 1).

Table 1. Stratification of projective rectification.

Input: Video frame i is available to read Image segmentation to identify the target building facade $\mathcal{F} \leftarrow$ Section 2.1 Line detection and segment clustering: $\mathbf{l}_v, \mathbf{l}_h$ in Part I of $\mathcal{F} \leftarrow$ Section 2.2	
Method using equally spaced parallel lines:	Method using vanishing points:
Part 1: Projective distortion removal 1. Group $\mathbf{l}_\lambda$ from $\mathbf{l}_h \leftarrow$ Section 2.3 2. $\mathbf{l}_\infty \leftarrow \mathbf{V} \leftarrow \mathbf{l}_\lambda$ in Section 2.3 3. $\mathbf{P} \leftarrow \mathbf{l}_\infty$ in Section 2.4.1 4. $\mathbf{x}' \leftarrow \mathbf{P}\mathbf{x}$ in $\mathbf{I}, \mathbf{l}'_z \leftarrow \mathbf{P}^{-T}\mathbf{l}_z, \mathbf{l}'_v \leftarrow \mathbf{P}^{-T}\mathbf{l}_v$	Projective distortion removal 1. Obtain vanishing points $\mathbf{v}_h, \mathbf{v}_z$ from $\mathbf{l}_h, \mathbf{l}_z \leftarrow$ Section 2.2 2. $\mathbf{l}_\infty = \mathbf{v}_h \times \mathbf{v}_z$ 3. $\mathbf{P} \leftarrow \mathbf{l}_\infty$ in Section 2.4.1 4. $\mathbf{x}' \leftarrow \mathbf{P}\mathbf{x}$ in $\mathbf{I}, \mathbf{v}'_h \leftarrow \mathbf{P}\mathbf{v}_h, \mathbf{v}'_z \leftarrow \mathbf{P}\mathbf{v}_z$
Part 2: Affine distortion correction 1. $\mathbf{A}_1 \leftarrow \theta \leftarrow \mathbf{l}'_z, \mathbf{l}'_v$ in Section 2.4.2 2. $\mathbf{A}_2 \leftarrow \mu \leftarrow$ Rectangular structures 2. $\mathbf{X} \approx \mathbf{A}_2\mathbf{A}_1\mathbf{x}'$ (with global scale factor)	Affine distortion correction 1. $\mathbf{A}_1 \leftarrow \theta \leftarrow \mathbf{v}'_h, \mathbf{v}'_z$ in Section 2.4.2 2. $\mathbf{A}_2 \leftarrow \mu \leftarrow$ Rectangular structures 2. $\mathbf{X} \approx \mathbf{A}_2\mathbf{A}_1\mathbf{x}'$ (with global scale factor)
Output: Displacements $\mathbf{d} = \Delta\mathbf{X}$, then go to the next video frame $i + 1$	

2.4.2. Affine Distortion Correction

Affine distortion correction makes angles in the rectified image equal to angles in the world. In this research, the affine distortion correction is accomplished in two steps, according to the following decomposition of the affine transformation $\mathbf{A}$

$$\mathbf{A} \approx \mathbf{A}_2 \mathbf{A}_1 = \begin{pmatrix} \mu & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & -\cot(\theta) & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (17)$$

where θ is the angle between the directions of $\mathbf{l}'_z$ and $\mathbf{l}'_h$, or say $\mathbf{v}'_h$ and $\mathbf{v}'_z$, and μ is viewed as the aspect ratio, which corresponds to the relative scale of the horizontal and vertical directions. In fact, the line-based information we have used to date only provides the constraint for θ in $\mathbf{A}_1$, while leaving the ambiguity caused by the unknown aspect ratio μ in $\mathbf{A}_2$. When we are most concerned with rectification of building facades—where rectangular structures always exist, such as the facade outline or windows—the aspect ratio μ can be easily acquired by the known width to height ratio of those rectangular structures. μ is also to be resolved from partial knowledge of the internal camera parameters in [42]. The above procedure of affine distortion correction is also described in detail in Table 1 (Part 2).

3. Experimental Case Studies

To experimentally validate the proposed method of vision-based displacement measurement, two case studies are presented in this section. In the first synthetic experimental test, the projective rectification was conducted directly based on the lines extracted from the regions of the target building itself. The appropriate regions to extract line segments for image rectification are discussed and verified for a typical high-rise building. In the second real-recorded case, line information was directly read from a reference board. We showed that while the scenes suffered, a strongly changing field of view taken by a hand-held camera, the proposed method still demonstrates a sub-pixel measurement accuracy.

3.1. Synthetic Experiments

3.1.1. A 30-Story Building Model

First, we evaluate the feasibility, accuracy and robustness of the proposed method using synthetic images, since no real recorded video of building structures subjected to earthquake ground motion was available at the present time. The Synthetic images were generated by four synthetic cameras observing a 3D finite element model of a 30-story building structure. The structure is a braced frame, with a height of 105 m, as shown in Figure 5. The layout and the member section of the structure are detailed in the figure. The load on the structure was modeled as seismic masses lumped at the floor levels. The finite element model of the structure was established in the ANSYS program [43], whereas the beam, column, and brace were simulated by the beam188 element, while the lumped mass was simulated by the mass21 element. Within each floor, the translational degrees of freedom of the nodes were coupled to comply the rigid floor assumption. The basic dynamic properties of the structure, including the natural period and the shape of the first six natural modes, are listed in Table 2. Specifically, the fundamental period $T_1 = 2.47$ s is in compliance with the engineering practice [44–46].

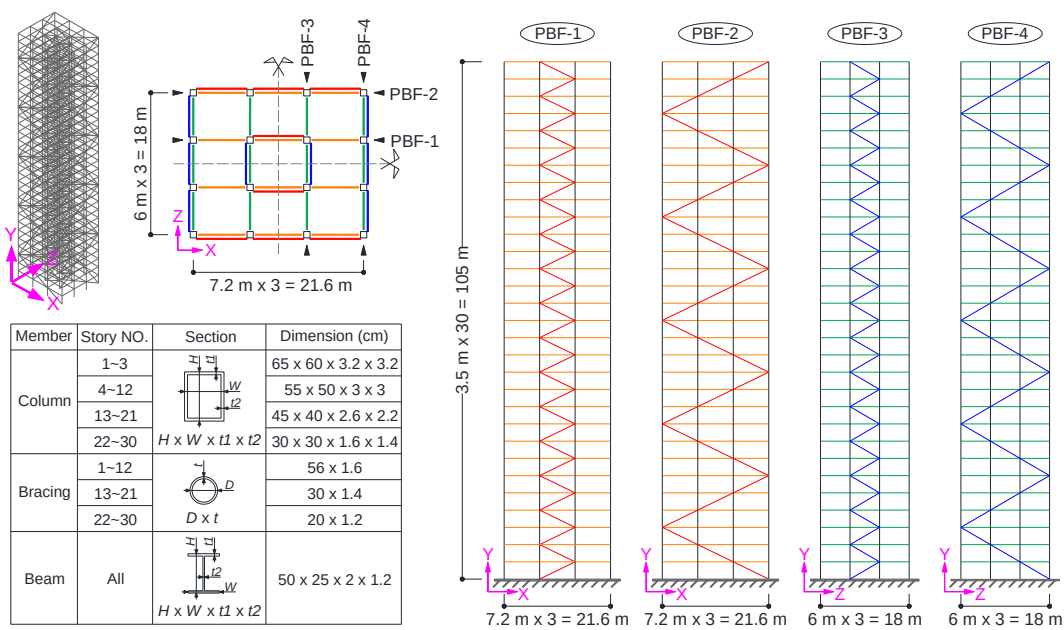


Figure 5. Basic information of the 30-story building model.

Table 2. Dynamic property of the building.

Order	Type	Period/s	Mode 1	Mode 3	Mode 5	Mode 2	Mode 4	Mode 6
1	Translational-Y	2.470						
2	Translational-X	2.251						
3	Translational-Y	0.736						
4	Translational-X	0.709						
5	Translational-Y	0.392						
6	Translational-X	0.387						

The structure was subjected to a ground acceleration along the X direction (see the world coordinate in Figure 5). The seismological information and the pseudo-acceleration spectrum of the excitation is shown in Figure 6a,b, respectively. The lateral displacement history of the roof is shown in Figure 6c. From Figure 6c, the peak lateral displacement of the roof is 0.3366m. The accordant story drift angle (ratio of the peak roof displacement to the structural height) is about 1/312 rad. Since there is a great deal of concern in the peak displacement response of the building during an earthquake event, only the corresponding structural displacements at $t = 25.86$ s were extracted and used to generate the synthetic images in this example.

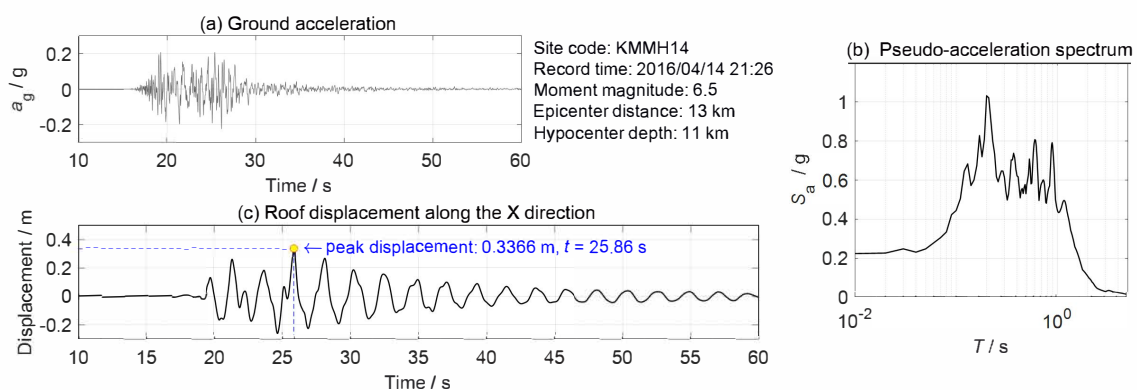


Figure 6. Earthquake excitation and displacement response of the building: (a) time history of the ground acceleration, (b) pseudo-acceleration spectrum, and (c) roof displacement.

3.1.2. Image Generation

Four artificially created views were generated from four separate synthetic cameras, as shown in Figures 7a and 8. For each view, a virtual pinhole camera with the image size of 1920×1080 pixels was modeled, with different camera positions and rotations. The camera was then adjusted by different focal lengths so that all line segments were in its field of view.

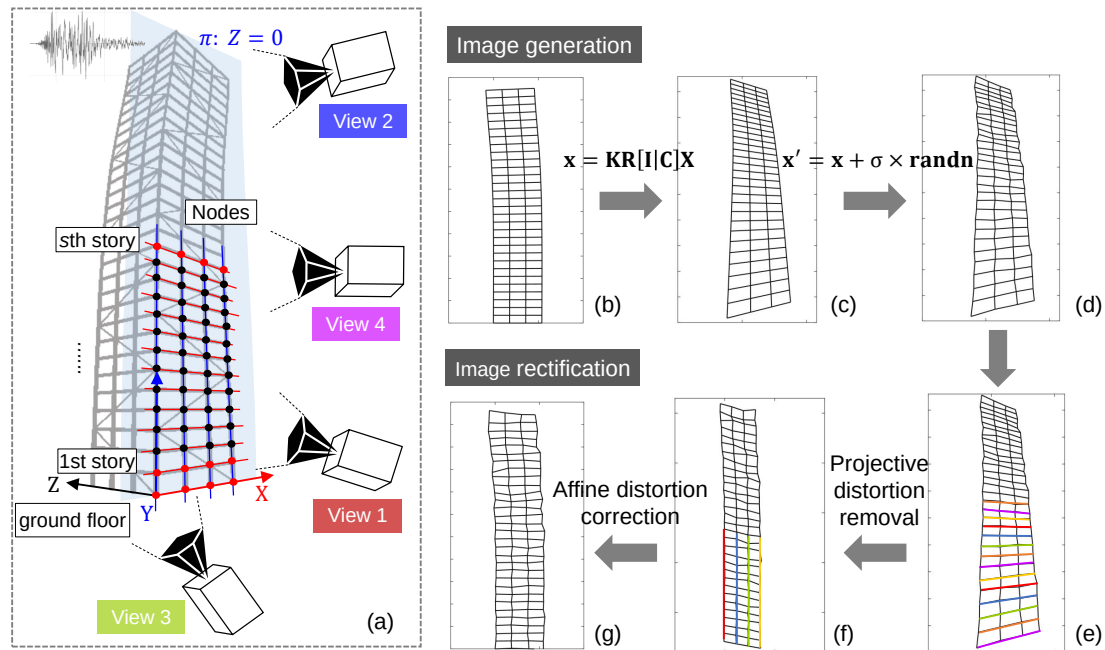


Figure 7. Processes of image generation and image rectification (building deformations are drawn to a scale of ten for enhanced visualizations): (a) positions of the camera, (b) world plane $Z=0$, (c) synthetic image after camera projection, (d) image perturbed with noise, (e) image with line detection, (f) image after projective distortion removal, (g) image after affine distortion correction.

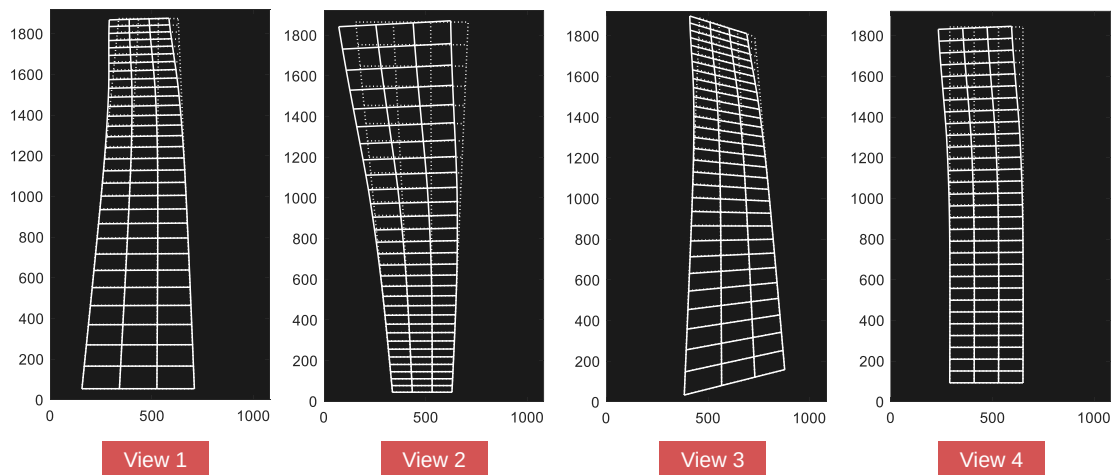


Figure 8. Synthetic views used in this example (building deformations are drawn to a scale of ten for enhanced visualizations).

The image coordinate system is assumed to be Euclidean coordinates with equal scales in both axial directions ($f_x = f_y = f$) and s is set to zero in this example. Details of the camera parameters for each synthetic view are listed in Table 3. It is noteworthy that the above camera parameters were only used for synthetic image generation. Displacement measurement using the proposed method requires no knowledge of camera parameters.

Table 3. Camera parameters of each view.

View	Extrinsic Parameters						Intrinsic Parameters		
	Camera Position (m)			Camera Rotation (rad)			Focal Length (pixel)	Principal Point Coordinate (pixel)	
	C_x	C_y	C_z	A	B	Γ	f	p_x	p_y
1	15	30	−30	−0.2	0	0	600	539.5	959.5
2	15	75	−30	0.2	0	0	600	539.5	959.5
3	0	30	−30	−0.2	0.2	0	600	539.5	959.5
4	15	52	−30	0	0	0	500	539.5	959.5

Line segments were generated by placing the nodes of beam elements at world plane $\pi : Z = 0$ as endpoints, see Figure 7a. The line segment $\mathbf{l}$ passing through two image endpoints $\mathbf{x}_1$ and $\mathbf{x}_2$ is then obtained by $\mathbf{l} = \mathbf{x}_1 \times \mathbf{x}_2$, where $\times$ is the vector product. If more than two points are provided, the method of orthogonal regression is used to estimate the line that gives the best fit to those points. The linear projection of 3D world line segments onto the image plane can also be achieved directly using Plücker coordinates. This procedure was described in detail in [47]. Note that only the nodes close to the ground floor with trivial displacements are limited for use in line generation and projective rectification. The vanishing line was estimated based on the equally spaced parallel lines generated by the nodes from ground floor to the s th story, $s = 8 - 22$.

To investigate the robustness of the proposed method, the coordinates of the endpoints in the image plane were additionally perturbed with independent and identically distributed Gaussian noise (Figure 7d), with a standard deviation of $\sigma = 0.5$ pixel and $\sigma = 1.0$ pixel, respectively. The relationship between the pixel coordinate and the physical dimension in the world system is about 0.040 m/pixel–0.074 m/pixel (bottom-top) in View 1. Monte Carlo simulations with 1000 trials for each noise level were performed under the above setup.

3.1.3. Measurement Results

We introduce the root mean square error (RMSE) between the image-estimated nodal coordinates in X-direction ($X_i, i = 1, \dots, N$) of the structure after image rectification and the corresponding ground truth coordinates (X_{Gi}) obtained by ANSYS in each image as follows

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (X_i - X_{Gi})^2} \quad (18)$$

The proposed method was evaluated and compared with the widely used, point-based image rectification technology, where the projective transformation $\mathbf{H}$ in Equation (5) is directly estimated by minimizing a robust maximum likelihood cost function using matched points [40]. Two cases, in which enough points (both black and red nodes in Figure 7a) and sparse points (only the red nodes in Figure 7a) were employed for projective rectification, were taken into consideration herein.

Results of the synthetic experiment are summarized in Figure 9 and typical line-based rectified images from each view with different noise levels are referred to Figure 10.

The RMSEs in Figure 9 were computed based on the overall nodes on each view. The results in the figures show that, when an appropriate region from the target building was used for projective rectification, the RMSE of measured displacements was below 10^{-3} m if no pixel noise exists in the image. For the cases with noises, the lowest RMSE was about 0.03 m for $\sigma = 0.5$ pixel, and 0.06 m for $\sigma = 1$ pixel, respectively. It is easily observed that there is an optimal value for the number of stories that can be used for projective rectification. The reason for this is that, on one hand, the more information (lines/points) is used for image rectification, the greater the accuracy that can be achieved. On the other hand, the displacements of the lines/points positioned at higher stories of the target

building were nontrivial under ground motion, and thus inappropriate to be involved in the projective rectification step. From the noise-free case, the measurement error with $RMSE = 37 \text{ mm}$ for $s = 22$ was shown to be 46 times larger than that with $RMSE = 0.8 \text{ mm}$ for $s = 14$. It appeared that the accuracy difference became less significant as the levels of noise went up. According to Figure 9, values of s , which led to the measurement errors that did not deviate much from the minimum, can range from 10 to 20 stories for both $\sigma = 0.5 \text{ pixel}$ and $\sigma = 1 \text{ pixel}$. It is thus proved that the abundant line information in the region from the ground to about $1/3 - 2/3$ height of the building itself might also be able to be used for projective rectification in this case. However, when the targeting building itself was used, the measurement results became very sensitive to the image noise. Measurement errors were 0.2% (0.0008 m/0.337 m) for no noise, 9% (0.03 m/0.337 m) for $\sigma = 0.5 \text{ pixel}$, and 18% (0.06 m/0.337 m) for $\sigma = 1 \text{ pixel}$. As we can see, when the pixel noise approaches one, the measurement results seem to be unacceptable. In such cases, a camera with higher resolution should be used, or a more strictly rigid region around the building should be applied for projective rectification.

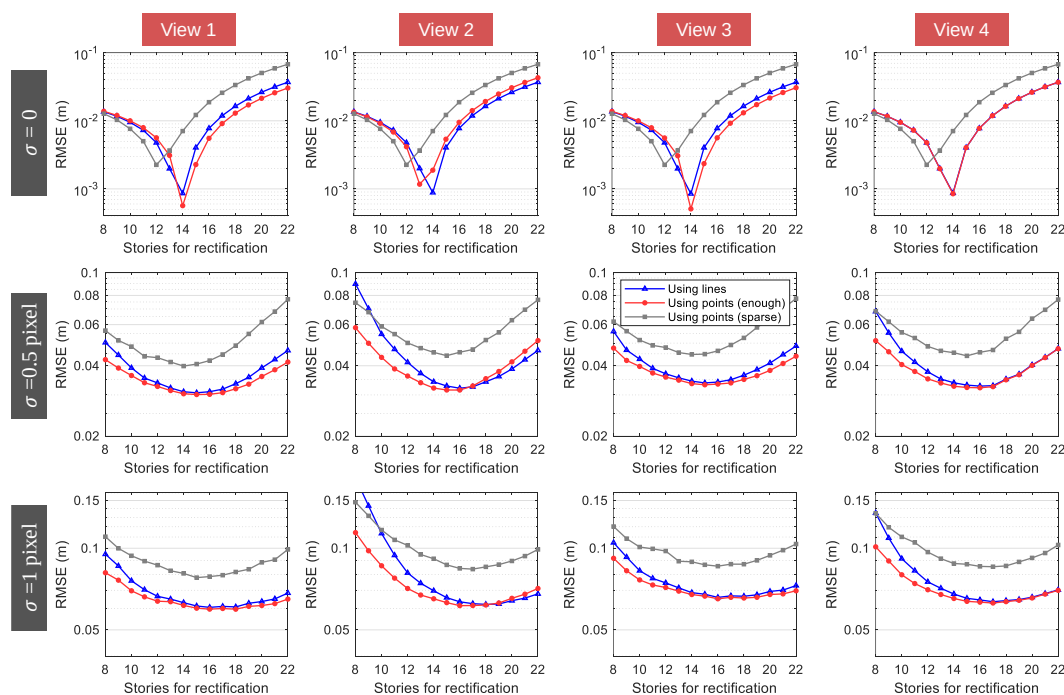


Figure 9. Measurement results for each view with different noise levels.

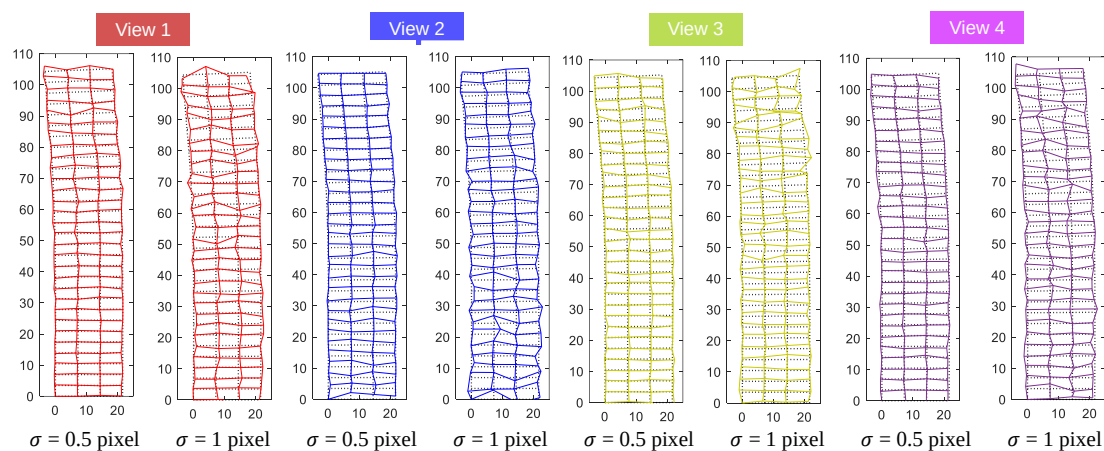


Figure 10. Rectified images from each view with different noise levels (Building deformations are drawn to a scale of ten only for enhanced visualizations).

Figure 9 also reports that our line-based method and the normally used, point-based method with the same number of points applied for image rectification behaved nearly identically, with the latter being slightly more accurate for small value of s . This was attributed to the accuracy degradation of the estimated orthogonal lines $\mathbf{l}_v$, which became more sensitive to the noise, with fewer points used for orthogonal regressions. Inaccurate $\mathbf{l}_v$ disturbed the precision of the affine distortion correction and resulted in errors to the final measurements in X-direction. A possible solution to circumvent this issue is to alternatively identify the normal $\mathbf{n}$ of the ground plane by plane detection techniques [48,49]. Since $\mathbf{n}$ is parallel to $\mathbf{l}_v$, it can be directly used for affine distortion correction in Section 2.4.2. This process will be discussed in detail in further research. Moreover, it should be pointed out that for low-texture scenes with man-made building facades in urban environments, line features are often more abundant and reliable than point features. If only sparse points were provided, measurement errors would significantly increase, as depicted in Figure 9 for the point-based rectification method, where the RMSE results change from the red lines to the gray lines. In such cases, using line features for displacement measurement is much more feasible and practicable.

For all of the four different views, the results of view 2 were the least accurate ones based on the overall RMSE depicted in Figure 9. This might be due to the fact that the corresponding camera was farthest from the region used for projective rectification among all of the synthetic views and this leads to a likely overall inaccuracy. If the RMSE for each story was calculated separately, view 2, on the other hand, resulted in the highest measurement accuracy for the upper region of the building, as shown in Figure 11. Figure 11 also shows that the measurement accuracy for the upper region decreased as the camera's position was down in the Y direction (refer to C_y in Table 3). That is to say, different regions of the target building had varying measurement accuracy for the image with perspective view, and the structural region nearest to the camera's position attained the greatest precision.

Attention should also be paid to view 4, since this view went through the least projective distortion but did not yield sufficiently accurate overall measurement results, as expected. The reason behind this is that focal length of this view had to be changed to 500 pixels so that all lines/points were in its field of view, while it was able to be set to 600 pixels for the other views. Based on the above observations, cameras are suggested to be positioned close to the ground/rigid regions of the structure if the accuracy of the overall motion is required, or near to the top area of the building if the roof displacement is more concerned, with powerful zoom lens and a large enough field of view for accurate measurement.

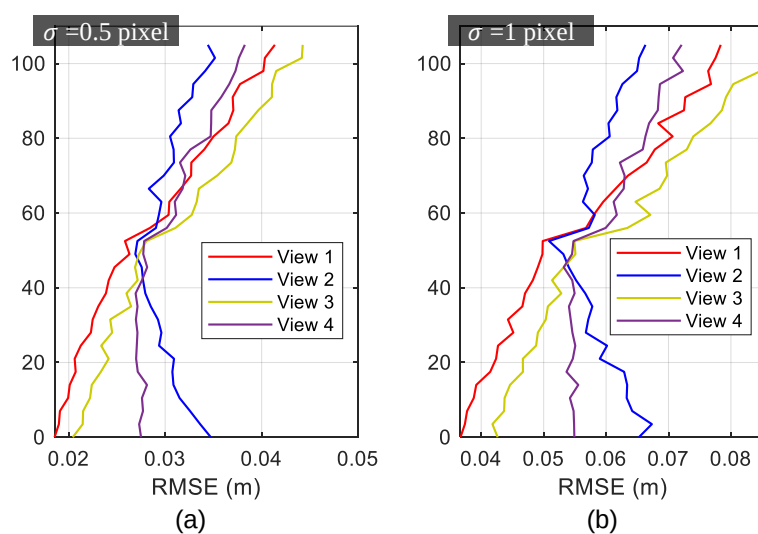


Figure 11. RMSE for each story: (a) $\sigma = 0.5$ pixel (b) $\sigma = 1$ pixel.

3.2. Experiments on Real Video Sequences

3.2.1. Experiment Test Setup

Since a high-rise tested structure was difficult to obtain due to the limitations of experimental facilities, in this experiment, we detected lines for projective rectification directly from a reference board instead of the target structure itself.

Figure 12a shows the overview of the experiment. The two-story base-isolated structure is with 140 mm width and 400 mm height. The masses of each story are about $m_0 = 2.100$ kg (base story), $m_1 = 3.247$ kg (first story) and $m_2 = 1.531$ kg (second story). The stiffness is about $k_1 = 2.3 \times 10^4$ N/m and $k_2 = 7.5 \times 10^3$ N/m for the first story and second story, respectively. In the base story, a BSG-H10 slider with frictional interfaces and tension springs with stiffness of $k_s = 2.3 \times 10^3$ N/m are assembled as a seismic isolator for this structure. The structure is bolted on the shaking table with the size of 400×400 mm. The scaled earthquake record of 2016 Kumamoto earthquake in Figure 12b is selected as the ground motion. The data of this record are from K-NET and KiK-net of Japanese NIED strong-motion seismograph networks [50].

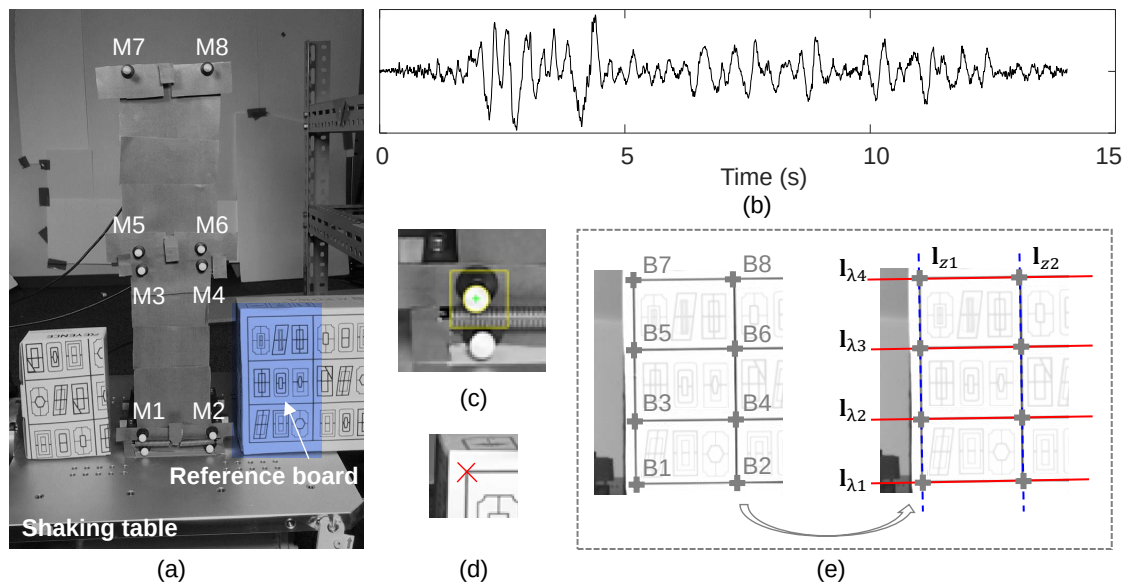


Figure 12. Experiment test setup: (a) overview of the two-story base-isolated structure with artificial tracking markers, (b) scaled earthquake ground motion acceleration record, (c) circular-based tracker, (d) correlation-based tracker, (e) reference board and lines extracted for projective rectification.

To validate the *relative displacement* results measured by the vision-based system, four laser transducers were installed on a shelf close to the shaking table as reference sensors. Absolute displacements and accelerations of the shaking table, the base story, the first story and the second story of the structure were additionally recorded by the reference laser transducers and accelerometers, respectively. Let a_g, a_0, a_1, a_2 denote the absolute accelerations of the shaking table, the base story, the first story and the second story of the structure, and $u_g^*, u_0^*, u_1^*, u_2^*$ denote the absolute displacement. It should be noted that the relative displacements u_0, u_1, u_2 with respect to the ground motion for each story were directly measured by the vision-based system, while they were reconstructed by the absolute displacements $u_g^*, u_0^*, u_1^*, u_2^*$ from the laser transducers via $u_0 = u_0^* - u_g^*, u_1 = u_1^* - u_g^*, u_2 = u_2^* - u_g^*$.

A hand-held iPhone 7 equipped with 28-mm lenses (12 MP, f/1.8, 1/3") was used for displacement measurement of the structure during the input of the seismic ground motion. Artificial white circular markers (M1-M8) glued to the base, first and second floor of the structure were tracked throughout the video by computing the coordinates of their gravity centers on the image plane (see Figure 12a,c).

Four equally spaced horizontal lines ($l_{\lambda 1} \sim l_{\lambda 4}$) and two vertical lines (l_{z1}, l_{z2}) were extracted by tracking the group of points (B1–B8) provided on the reference board, as shown in Figure 12e. B1–B8 were tracked based on the correlation-based template matching method. The method using equally spaced parallel lines to estimate the vanishing line was applied, since available vertical lines were limited.

The iPhone was subjected to large motions, such as translations and rotations, while the structure was subjected to the seismic ground motion. Tracking results of the markers without any image rectification are shown in Figure 13, taking the trajectory of the second story (the mean value of M7–M8's tracking coordinates (in pixel)) as an example.

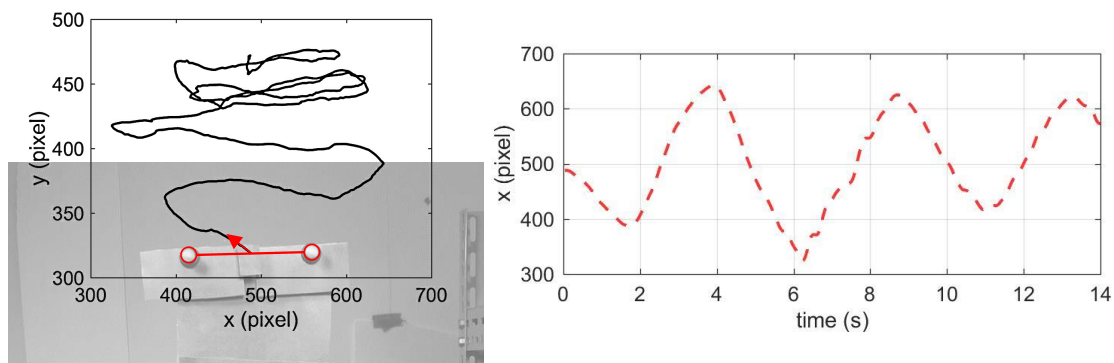


Figure 13. Trajectory of the markers in the second story before image rectification.

As is seen, real displacements of the structure were overwhelmed by the strongly changing field of view caused by the moving camera and seemed hard to distinguish from their overall trajectories if no image rectification was involved.

3.2.2. Measurement Results

The final vision-based displacement measurement results using the proposed line-based projective rectification method are graphically shown in Figure 14.

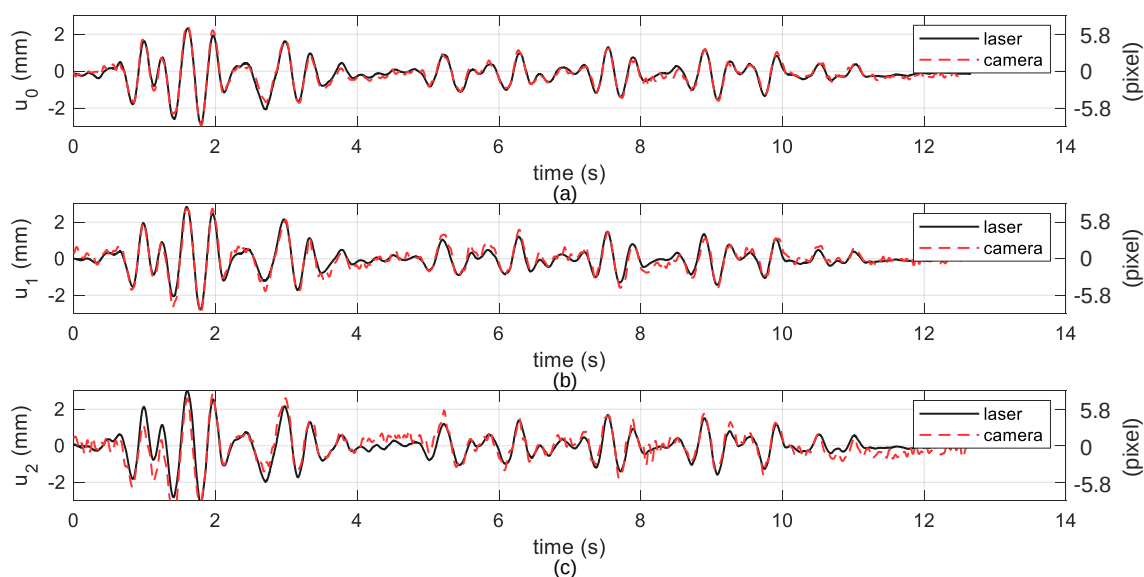


Figure 14. Measured results from the proposed method: (a) the base story displacement u_0 , (b) the first story displacement u_1 , (c) the second story displacement u_2 .

It is clear that the proposed method provides the measurement results with comparable accuracy to that obtained by referenced laser transducers. Next, the accuracy of the displacement measurement was additionally assessed by computing the nonlinear frictional force on the base interfaces of the structure from the fusion of measured displacement data and acceleration data using the following equation

$$F = -(a_1 m_1 + a_2 m_2 + a_0 m_o) - k_s u_0 \quad (19)$$

where a_1, a_2 and a_0 are the measured accelerations of the first, second and base story, respectively. Figure 15 demonstrates that the vision-based measurement results agree well with the laser-based results in system force estimation. With the aid of successfully measured displacement data, more key physical performances of the mechanical system can be investigated and the perdition of system's further behavior for prognosis in structural health monitoring is thus able to be delivered.

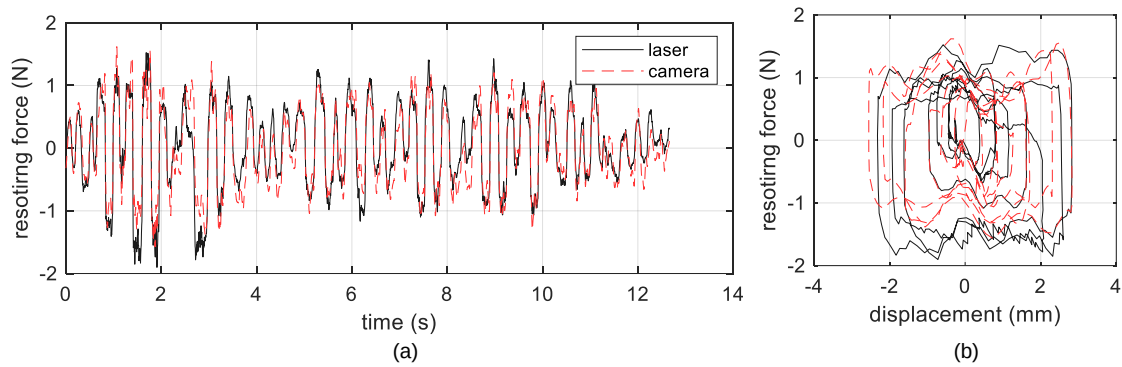


Figure 15. Frictional force F computed by the measured displacement and acceleration data: (a) time history of F , (b) u_0 vs. F .

To further validate the effectiveness of the proposed method, the displacement measurement results obtained by the point-based projective rectification method, in which the same points B1-B8 were specified to be used to compute the maximum likelihood estimation of $\mathbf{H}$ in Equation (5), are presented in Figure 16.

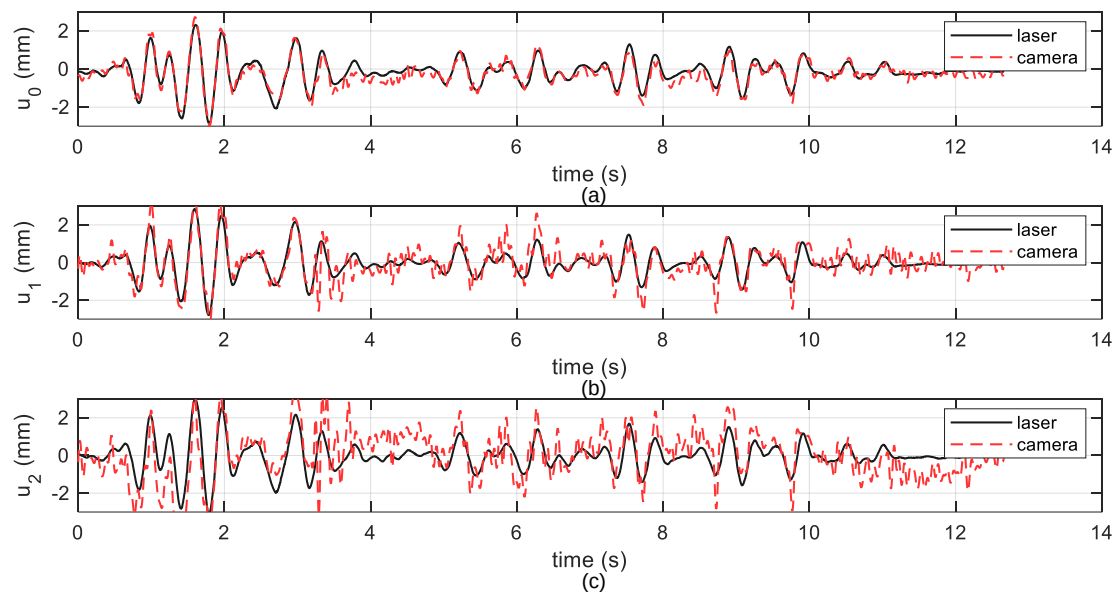


Figure 16. Measured results based on point-based projective rectification method: (a) the base story displacement u_0 , (b) the first story displacement u_1 , (c) the second story displacement u_2 .

Comparing the results from different methods for projective rectification, the line-based method has a much better performance in terms of algorithm accuracy. In theory, the line-based rectification and point-based rectification should be equivalent, as they were provided with the same information. Nevertheless, it is found that the line-fitting is generally more noise-resistant than merely point detection, especially for the cases when feature information is limited. For a complete exposition of this inference and special attentions which should be paid when using the line-based method, the reader is referred to [51].

Finally, we computed the RMSE between the vision-based displacement measurements and the laser references. The RMSE is plotted for each video frame in pixel.

According to Figure 17, it is seen that errors in most frames of our line-based method are below one pixel. The average RMSE is 0.68 pixel per frame while 1.25 pixel for the point-based method. The measure error is about 8% (0.68 pixel/8.08 pixel) for the proposed approach, which is satisfied and acceptable for dynamic displacement measure. Comparing the tracking history of RMSE also allows us to quantitatively assess the contribution of the matrix V update strategy by the iterative minimization techniques proposed in Section 2.3. When we re-ran the algorithm without the update of V using only the value estimated by the DLT algorithm, the resulting average error increased to 1.41 pixel per frame. These results demonstrate that the accurate estimation of the matrix V also provides an apparent boost to measuring performance.

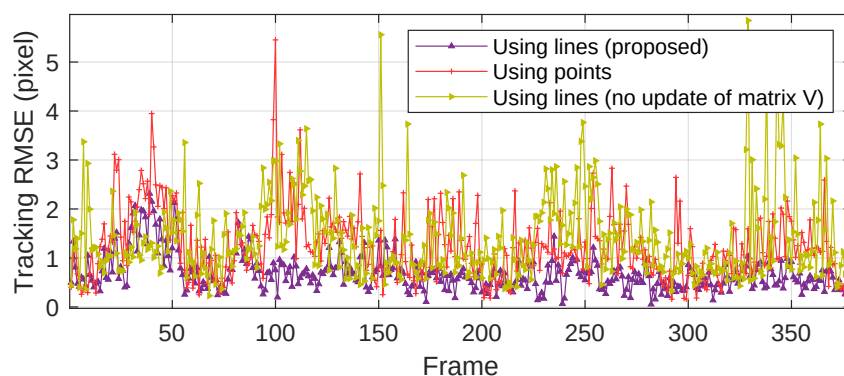


Figure 17. Tracking RMSE for each frame.

4. Conclusions and Discussion

A new vision-based system for seismic displacement measurement of building structures is developed in this research. The technique only requires line-based geometric relationships on the building facade plane to be extracted from a single, uncalibrated, perspective view. To remove the projective/affine distortion in the original image, stratification of the projective rectification technique, which is accomplished by employing parallelism and orthogonality of lines to compute the projective and affine components of the transformation sequentially, is introduced in this research. Synthetic examples and experiments on real video sequences are explored. The accuracy of displacement estimates and the robustness to image noise of the proposed method were validated.

To date, we have shown some advantages of the proposed method in projective rectification and structural displacement measurement. In practice, however, several attentions should be paid and some challenges still remain open:

- Only step (3)–(5) in Section 2 were validated via experimental case studies in this study, since no real recorded video for seismic-induced motion measurement of building structures was available at the present time. A vision-based system with the newly released Canon EOS R5 camera (4 K at 120 fps) has already been incorporated into a structural health monitoring system of a high-rise building and the proposed method is expected to contribute to the further research;

- Although sub-pixel level accuracy was attained in this study, the real application of this image-processing technique might be inferior to the laboratory precision since, in the real world, there exists not only pixel noise, but also image distortion as well as line segment extraction error, which make the problem much more challenging;
- When using vanishing points to rectify the image, any horizontal/vertical parallel lines, coplanar or not, can be involved in the algorithm, while the method using equal spaced lines requires coplanar parallel lines;
- When structures are subjected to out-of-plane motion, the proposed method is still applicable. To measure in three dimensions, image rectifications with respect to two mutually orthogonal planes of the building, e.g., $Z = 0$ and $X = 0$ in Figure 7a, should be employed from a single image. Therefore, three dimensional measurements may incur the trade-off problem between measurement resolution and the field of view. A higher resolution, such as 4 K (e.g., 3840×2160), is suggested to be set for accurate measurement.

Author Contributions: Conceptualization, J.G. and I.T.; methodology, J.G.; software, Y.X.; resources, K.F.; writing—original draft preparation, J.G.; writing—review and editing, Y.X.; visualization, K.F.; supervision, I.T. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: The authors thank the whole joint research team at Kyoto University, in particular Miss Jian Jiao for providing all the experimental data shown here. The earthquake record from K-NET and KiK-net of Japanese NIED strong motion seismograph networks was used.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Takewaki, I.; Fujita, K.; Yoshitomi, S. Uncertainties in long-period ground motion and its impact on building structural design: case study of the 2011 Tohoku (Japan) earthquake. *Eng. Struct.* **2013**, *49*, 119–134. [\[CrossRef\]](#)
2. Kasai, K.; Pu, W.; Wada, A. Responses of controlled tall buildings in Tokyo subjected to the Great East Japan earthquake. In Proceedings of the International Symposium on Engineering Lessons Learned from the 2011 Great East Japan earthquake, Tokyo, Japan, 1–4 March 2012; pp. 1–4.
3. Stiros, S.C. Errors in velocities and displacements deduced from accelerographs: An approach based on the theory of error propagation. *Soil Dyn. Earthq. Eng.* **2008**, *28*, 415–420. [\[CrossRef\]](#)
4. Herring, T.; Gu, C.; Toksöz, M.N.; Parol, J.; Al-Enezi, A.; Al-Jeri, F.; Al-Qazweeni, J.; Kamal, H.; Büyükoztürk, O. GPS measured response of a tall building due to a distant Mw 7.3 earthquake. *Seismol. Res. Lett.* **2019**, *90*, 149–159. [\[CrossRef\]](#)
5. Nickitopoulou, A.; Protopsalti, K.; Stiros, S. Monitoring dynamic and quasi-static deformations of large flexible engineering structures with GPS: Accuracy, limitations and promises. *Eng. Struct.* **2006**, *28*, 1471–1482. [\[CrossRef\]](#)
6. Stanbridge, A.; Ewins, D. Modal testing using a scanning laser Doppler vibrometer. *Mech. Syst. Signal Process.* **1999**, *13*, 255–270. [\[CrossRef\]](#)
7. Choi, I.; Kim, J.; Kim, D. A target-less vision-based displacement sensor based on image convex hull optimization for measuring the dynamic response of building structures. *Sensors* **2016**, *16*, 2085. [\[CrossRef\]](#) [\[PubMed\]](#)
8. Zhang, D.; Guo, J.; Lei, X.; Zhu, C. A high-speed vision-based sensor for dynamic vibration analysis using fast motion extraction algorithms. *Sensors* **2016**, *16*, 572. [\[CrossRef\]](#)
9. Feng, D.; Feng, M.Q.; Ozer, E.; Fukuda, Y. A vision-based sensor for noncontact structural displacement measurement. *Sensors* **2015**, *15*, 16557–16575. [\[CrossRef\]](#) [\[PubMed\]](#)
10. Yoon, H.; Shin, J.; Spencer Jr, B.F. Structural displacement measurement using an unmanned aerial system. *Comput.-Aided Civil Infrastruct. Eng.* **2018**, *33*, 183–192. [\[CrossRef\]](#)
11. Yoneyama, S.; Ueda, H. Bridge deflection measurement using digital image correlation with camera movement correction. *Mater. Trans.* **2012**, *53*, 285–290. [\[CrossRef\]](#)

12. Dworakowski, Z.; Kohut, P.; Gallina, A.; Holak, K.; Uhl, T. Vision-based algorithms for damage detection and localization in structural health monitoring. *Struct. Control Health Monit.* **2016**, *23*, 35–50. [\[CrossRef\]](#)
13. Yoon, H.; Elanwar, H.; Choi, H.; Golparvar-Fard, M.; Spencer Jr, B.F. Target-free approach for vision-based structural system identification using consumer-grade cameras. *Struct. Control Health Monit.* **2016**, *23*, 1405–1416. [\[CrossRef\]](#)
14. Chen, J.G.; Davis, A.; Wadhwa, N.; Durand, F.; Freeman, W.T.; Büyüköztürk, O. Video camera-based vibration measurement for civil infrastructure applications. *J. Infrastruct. Syst.* **2017**, *23*, B4016013. [\[CrossRef\]](#)
15. Cheng, C.; Kawaguchi, K. A preliminary study on the response of steel structures using surveillance camera image with vision-based method during the Great East Japan Earthquake. *Measurement* **2015**, *62*, 142–148. [\[CrossRef\]](#)
16. Kim, S.W.; Kim, N.S. Dynamic characteristics of suspension bridge hanger cables using digital image processing. *NDT E Int.* **2013**, *59*, 25–33. [\[CrossRef\]](#)
17. Szeliski, R. *Computer Vision: Algorithms and Applications*; Springer Science & Business Media: Berlin, Germany, 2010.
18. Wildenauer, H.; Hanbury, A. Robust camera self-calibration from monocular images of Manhattan worlds. In Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition, Rhode Island, RI, USA, 16–21 June 2012; pp. 2831–2838.
19. Xu, C.; Zhang, L.; Cheng, L.; Koch, R. Pose estimation from line correspondences: A complete analysis and a series of solutions. *IEEE Trans. Pattern Anal. Mach. Intell.* **2016**, *39*, 1209–1222. [\[CrossRef\]](#)
20. Topal, C.; Akinlar, C. Edge drawing: A combined real-time edge and segment detector. *J. Vis. Commun. Image Represent.* **2012**, *23*, 862–872. [\[CrossRef\]](#)
21. Von Gioi, R.G.; Jakubowicz, J.; Morel, J.M.; Randall, G. LSD: A fast line segment detector with a false detection control. *IEEE Trans. Pattern Anal. Mach. Intell.* **2008**, *32*, 722–732. [\[CrossRef\]](#)
22. López, J.; Santos, R.; Fdez-Vidal, X.R.; Pardo, X.M. Two-view line matching algorithm based on context and appearance in low-textured images. *Pattern Recognit.* **2015**, *48*, 2164–2184. [\[CrossRef\]](#)
23. Xu, Y.; Oh, S.; Hoogs, A. A minimum error vanishing point detection approach for uncalibrated monocular images of man-made environments. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Portland, OR, USA, 25–27 June 2013; pp. 1376–1383.
24. Wu, L.J.; Casciati, F.; Casciati, S. Dynamic testing of a laboratory model via vision-based sensing. *Eng. Struct.* **2014**, *60*, 113–125. [\[CrossRef\]](#)
25. Guo, J.; Jiao, J.; Fujita, K.; Takewaki, I. Damage identification for frame structures using vision-based measurement. *Eng. Struct.* **2019**, *199*, 109634. [\[CrossRef\]](#)
26. Otsu, N. A threshold selection method from gray-level histograms. *IEEE T. Syst. Man. Cy.-S.* **1979**, *9*, 62–66. [\[CrossRef\]](#)
27. Dhanachandra, N.; Manglem, K.; Chanu, Y.J. Image segmentation using K-means clustering algorithm and subtractive clustering algorithm. *Procedia Comput. Sci.* **2015**, *54*, 764–771. [\[CrossRef\]](#)
28. Minaee, S.; Boykov, Y.; Porikli, F.; Plaza, A.; Kehtarnavaz, N.; Terzopoulos, D. Image segmentation using deep learning: A survey. *arXiv* **2020**, arXiv:2001.05566.
29. Teboul, O.; Simon, L.; Koutsourakis, P.; Paragios, N. Segmentation of building facades using procedural shape priors. In Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, CA, USA, 13–18 June 2010; pp. 3105–3112.
30. Hernández, J.; Marcotegui, B. Morphological segmentation of building façade images. In Proceedings of the 2009 16th IEEE International Conference on Image Processing (ICIP), Cairo, Egypt, 7–10 November 2009; pp. 4029–4032.
31. Wendel, A.; Donoser, M.; Bischof, H. Unsupervised facade segmentation using repetitive patterns. In *Joint Pattern Recognition Symposium*; Springer: Berlin, Germany, 2010; pp. 51–60.
32. Canny, J. A computational approach to edge detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **1986**, 679–698. [\[CrossRef\]](#)
33. Duda, R.O.; Hart, P.E. Use of the Hough transformation to detect lines and curves in pictures. *Commun. ACM* **1972**, *15*, 11–15. [\[CrossRef\]](#)
34. Akinlar, C.; Topal, C. EDLines: A real-time line segment detector with a false detection control. *Pattern Recognit. Lett.* **2011**, *32*, 1633–1642. [\[CrossRef\]](#)

35. Von Gioi, R.G.; Jakubowicz, J.; Morel, J.M.; Randall, G. LSD: a line segment detector. *Image Process. Line* **2012**, *2*, 35–55. [[CrossRef](#)]
36. Schaffalitzky, F.; Zisserman, A. Planar grouping for automatic detection of vanishing lines and points. *Image Vis. Comput.* **2000**, *18*, 647–658. [[CrossRef](#)]
37. Lezama, J.; Grompone von Gioi, R.; Randall, G.; Morel, J.M. Finding vanishing points via point alignments in image primal and dual domains. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Ohio, OH, USA, 24–27 June 2014; pp. 509–515.
38. Almansa, A.; Desolneux, A.; Vamech, S. Vanishing point detection without any a priori information. *IEEE Trans. Pattern Anal. Mach. Intell.* **2003**, *25*, 502–507. [[CrossRef](#)]
39. Horn, R.A.; Horn, R.A.; Johnson, C.R. *Topics in Matrix Analysis*; Cambridge University Press: Cambridge, UK, 1994.
40. Hartley, R.; Zisserman, A. *Multiple View Geometry in Computer Vision*; Cambridge University Press: Cambridge, UK, 2003.
41. Liebowitz, D.; Zisserman, A. Metric rectification for perspective images of planes. In Proceedings of the 1998 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No. 98CB36231), Santa Barbara, CA, USA, 25 June 1998; pp. 482–488.
42. Liebowitz, D.; Criminisi, A.; Zisserman, A. *Creating Architectural Models from Images*. *Computer Graphics Forum*; Blackwell Publishers Ltd.: Oxford, UK; Boston, MA, USA, 1999; Volume 18, pp. 39–50. Available online: <https://doi.org/10.1111/1467-8659.00326> (accessed on 1 May 2020).
43. *Multiphysics*, A Version 16.0; ANSYS, Inc.: Canonsburg, PA, USA, 2015.
44. Satake, N.; Suda, K.i.; Arakawa, T.; Sasaki, A.; Tamura, Y. Damping evaluation using full-scale data of buildings in Japan. *Int. J. Struct. Eng.* **2003**, *129*, 470–477. [[CrossRef](#)]
45. Tremblay, R. Fundamental periods of vibration of braced steel frames for seismic design. *Earthq. Spectra.* **2005**, *21*, 833–860. [[CrossRef](#)]
46. Kwon, O.S.; Kim, E.S. Evaluation of building period formulas for seismic design. *Earthq. Eng. Struct. Dyn.* **2010**, *39*, 1569–1583. [[CrossRef](#)]
47. Bartoli, A.; Sturm, P. Structure-from-motion using lines: Representation, triangulation, and bundle adjustment. *Comput. Vis. Image Underst.* **2005**, *100*, 416–441. [[CrossRef](#)]
48. Zhou, J.; Li, B. Homography-based ground detection for a mobile robot platform using a single camera. In Proceedings of the 2006 IEEE International Conference on Robotics and Automation, Orlando, FL, USA, 15–19 May 2006; pp. 4100–4105.
49. Liu, C.; Kim, K.; Gu, J.; Furukawa, Y.; Kautz, J. Planercnn: 3D plane detection and reconstruction from a single image. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 16–20 June 2019; pp. 4450–4459.
50. Aoi, S.; Kunugi, T.; Fujiwara, H. Strong-motion seismograph network operated by NIED: K-NET and KiK-net. *J. Jpn. Assoc. Earthq. Eng.* **2004**, *4*, 65–74. [[CrossRef](#)]
51. Zeng, H.; Deng, X.; Hu, Z. A new normalized method on line-based homography estimation. *Pattern Recognit. Lett.* **2008**, *29*, 1236–1244. [[CrossRef](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).