13th International Conference on
Structural Health Monitoring of Intelligent Infrastructure
DOI: 10.3217/978-3-99161-057-1-057

Machine Vision-Based Super-Resolution Reconstruction for High-Precision Displacement Monitoring of Hydraulic Structures

You Yang[1,2], Bo chen[1,2], Weiqi Liu[3], Zekai Ma[1,2]
[1]The National Key Laboratory of Water Disaster Prevention, Hohai University, Nanjing 210098, China;
[2]College of Water Conservancy and Hydropower Engineering, Hohai University, Nanjing 210098, China;
[3]POWERCHINA Chengdu Engineering Corporation Limited, Chendu 610072, China
email: 1638615936@qq.com, chenbo@hhu.edu.cn, sdgmlwq@163.com, 764987173@qq.com

ABSTRACT: In response to the issues of high cost, limited monitoring accuracy, and susceptibility to environmental factors in traditional hydraulic structure displacement automation monitoring methods, a non-contact intelligent monitoring method based on machine vision image super-resolution reconstruction is proposed. This method uses artificial targets as markers and combines a high-order image degradation model with a camera to analyze real monitoring scenarios, carry out image data collection, and perform displacement calculation. It innovatively introduces a feature fusion attention mechanism to improve the Real-ESRGAN network and generator, enabling the reconstruction of image contours and fine details to enhance displacement calculation accuracy. Laboratory and field test results show that this method can effectively improve image resolution and clarity, achieving sub-pixel and millimeter-level precise monitoring of hydraulic structure surface displacement. Compared with traditional super-resolution algorithms and target tracking methods, the improved Real-ESRGAN algorithm performs the best, with a coefficient of determination (R^2) of up to 0.9975, an average absolute error (MAE) as low as 0.5552, and residual errors controlled within 5mm. The edge contours and details in the images are successfully reconstructed, effectively improving the displacement monitoring accuracy of hydraulic structures based on machine vision.

KEY WORDS: Machine vision displacement monitoring; Image super-resolution reconstruction; Real-ESRGAN improvement; Feature fusion attention mechanism; Hydraulic structure safety monitoring.

1 INTRODUCTION

As critical infrastructure in water resource management systems, hydraulic engineering projects play a pivotal role in ensuring water security and sustaining watershed economic development. Efficient and stable operation of these structures not only guarantees essential water supply for socio-economic sustainability but also serves as a vital safeguard for regional ecological security [1,2]. In the field of structural health monitoring (SHM), accurate surface displacement measurement forms the basis for safety assessment and early warning systems [3]. However, existing monitoring technologies exhibit notable limitations: traditional displacement measurement methods are not only costly but also susceptible to environmental interference, while demonstrating inadequate responsiveness to sudden structural risks [4-6].

A further complication arises from the multi-physics coupling effects (including hydraulic loads, thermal stresses, and mechanical vibrations) that hydraulic structures endure during service. These complex interactions induce nonlinear deformation behaviors, potentially leading to progressive damage or even catastrophic failure, posing dual threats to both structural integrity and economic viability [7]. Consequently, there is an urgent need to develop intelligent, automated displacement monitoring methods and establish digitalized smart monitoring systems to enhance lifecycle safety management of hydraulic structures.

Current displacement monitoring techniques for hydraulic structures can be broadly categorized into contact-based and non-contact approaches [8]. Contact-based methods, such as strain gauges and fiber-optic sensors, suffer from complex installation and environmental sensitivity, making them unsuitable for long-term monitoring [9, 10].

In the field of non-contact structural monitoring, laser displacement sensors[11][12], machine vision[13][14], total stations[15][16], and Global Navigation Satellite Systems (GNSS) [17][18] have emerged as critical technologies for deformation monitoring of large-scale hydraulic structures such as sluice gates and dams, owing to their non-invasive nature and operational flexibility.

Laser displacement sensors employ optical triangulation principles, utilizing photoelectric receivers including position-sensitive detectors (PSD), charge-coupled devices (CCD), and complementary metal-oxide-semiconductor (CMOS) sensors to achieve high-precision displacement measurements. However, the measurement accuracy of this technology exhibits an inverse correlation with monitoring distance due to the inherent limitations of laser reflection-based triangulation, significantly reducing its suitability for long-range monitoring scenarios. Furthermore, the installation requirements for reflective targets impose additional constraints on deployment flexibility, particularly for vertical displacement monitoring of hydraulic structures[19].

GNSS technology has demonstrated robust capabilities in dynamic displacement monitoring for structural health monitoring (SHM) and seismic engineering applications. Nevertheless, its sampling frequency is constrained by satellite signal update rates, limiting its effectiveness in high-frequency vibration monitoring. Compared to conventional sensors such as accelerometers, GNSS also exhibits inferior measurement accuracy in the vertical direction[20].

Total stations, as multifunctional surveying instruments, integrate high-precision distance measurement, angular measurement, coordinate acquisition, and elevation determination, making them widely applicable in engineering surveying and structural monitoring[15][16][21]. However, in

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

practical applications for dam displacement monitoring, these instruments face challenges such as unfavorable cost-benefit ratios, line-of-sight obstructions at measurement points, and operational complexity, which hinder their widespread adoption in large-scale engineering projects.

To address these challenges, machine vision-based displacement monitoring enhanced by image super-resolution reconstruction has emerged as a promising solution. By employing deep learning algorithms to recover fine details from low-resolution images, this method overcomes the precision constraints of traditional vision-based measurements while avoiding the range limitations of laser displacement sensors and the environmental dependencies of drone-based remote sensing. Compared to fiber-optic sensing and conventional SHM systems, this approach offers superior flexibility, cost efficiency, and adaptability, providing a novel pathway for high-precision, automated displacement monitoring of hydraulic structures.

Recent advances in computer vision have demonstrated its potential in SHM applications. such as the Harris corner detection method[22], the Lucas-Kanade (LK) optical flow matching algorithm[23], and template matching algorithms[24] are employed.For instance, Yoon et al. [23] combined Harris corner detection with the Lucas-Kanade (LK) optical flow algorithm to achieve high-precision displacement tracking in building structures. Brownjohn et al.[25] deployed a vision-based monitoring system on the Humber Bridge, validating its engineering applicability. Bocian et al. [26] extracted modal parameters of cable-stayed bridges using template matching and sparse LK optical flow, enabling vibration characteristic analysis. Kohut et al.[27] applied digital image correlation (DIC) to measure structural deflection, providing intuitive deformation indicators.

Algorithmic optimizations have further enhanced performance: Wu et al. [28] improved computational efficiency in template matching for real-time monitoring; Molina-Viedma[29] integrated phase-based motion magnification with digital image processing techniques [30] to identify the modal shapes of a cantilever beam, thereby overcoming the limitation of conventional digital image methods in accurately capturing high-frequency structural vibrations. Khuc et al. [22] refined Harris corner extraction with FREAK descriptors for sub-pixel matching accuracy. Guo et al. [31] leveraged projection correction to enhance displacement measurement under seismic conditions.

Despite these advancements, machine vision-based monitoring for hydraulic structures—particularly gates and dams—remains underdeveloped due to complex operational environments and stringent safety requirements. Existing systems still face data acquisition constraints, lacking a mature framework for deformation monitoring.

This study proposes a non-contact intelligent monitoring method integrating machine vision with image super-resolution reconstruction to address the cost, precision, and environmental limitations of conventional techniques. The methodology employs artificial targets as fiducial markers, coupled with high-resolution imaging and higher-order image degradation modeling, to achieve precise displacement data acquisition. At the algorithmic level, we innovatively introduce a feature fusion attention mechanism to enhance the Real-ESRGAN network, specifically improving its contour detail reconstruction and texture recovery capabilities. This approach is expected to significantly elevate displacement calculation accuracy, offering a cost-effective, intelligent solution for hydraulic structure safety monitoring.

2 DEEP LEARNING-BASED SUPER-RESOLUTION METHODS: OPTIMIZATION AND APPLICATION OF GENERATIVE ADVERSARIAL NETWORKS

This study aims to transcend the limitations of conventional image resolution through deep learning techniques, thereby providing higher-precision machine vision measurement solutions for structural displacement monitoring in engineering applications.The following sections will focus on image super-resolution algorithms based on Generative Adversarial Networks (GANs), offering in-depth analysis of their core principles, network architectures, and optimization strategies.

2.1 Image Super-Resolution Based on Generative Adversarial Networks

The Super-Resolution Generative Adversarial Network (SRGAN) architecture primarily consists of two core modules engaged in adversarial competition: the Generator and the Discriminator, as illustrated in Figure 1. During model training, low-resolution (LR) images serve as input data to the generator network, which performs nonlinear transformations through deep neural networks to produce high-resolution (HR) images. Subsequently, both the generated super-resolution images and authentic HR images form the input sample space for the discriminator, which evaluates the authenticity probability of input samples through feature extraction and pattern recognition.Within this adversarial training framework, the generator optimizes its parameter space to minimize the distribution divergence between generated and real samples, aiming to produce super-resolution images with high visual fidelity. Concurrently, the discriminator continuously enhances its discriminative capability to maximize identification accuracy of generated samples. This minimax game process achieves dynamic equilibrium that ultimately enables the generator to produce super-resolution reconstruction results perceptually indistinguishable from genuine high-resolution images.

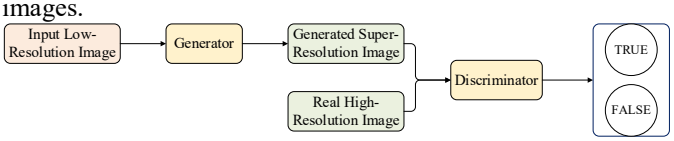


Figure 1. Architectural illustration of the Generative Adversarial Network framework

2.2 Image Super-Resolution Based on Generative Adversarial Networks

The Enhanced Super-Resolution Generative Adversarial Network (ESRGAN) represents an advanced generative adversarial network for super-resolution that achieved state-of-the-art performance in the field of image super-resolution at its time of introduction. This method primarily builds upon the aforementioned SRGAN architecture while implementing several key improvements.

To enhance model stability and generalization capability while reducing computational complexity, ESRGAN eliminates all

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

batch normalization layers from the original SRGAN framework and replaces the basic residual blocks with Residual-in-Residual Dense Blocks (RRDBs). Each RRDB consists of three dense blocks, with each dense block comprising five convolutional layers. The RRDB architecture effectively combines the advantages of dense connections and multi-level residual networks while removing block normalization, thereby reducing computational overhead while simultaneously minimizing artifacts in generated images.

ESRGAN further modifies the loss function to produce more realistic super-resolution outputs through two principal innovations: (1) replacing the original feature extractor with a VGG16 network, and (2) introducing a Relativistic average Discriminator (RaD) to substitute the conventional discriminator architecture.The BN-free residual block and Residual-in-Residual Dense Block (RRDB) structure are shown in Figure 2.

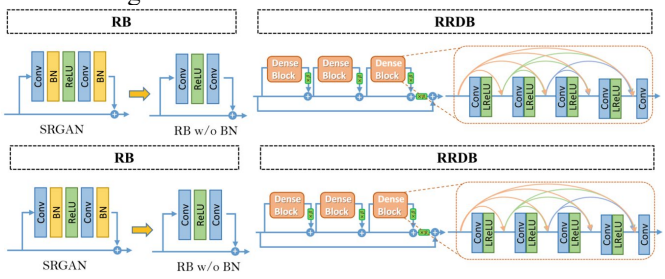


Figure 2. The BN-free residual block and Residual-in-Residual Dense Block (RRDB) structure

2.3 *High-Order Image Degradation Model*

This study innovatively proposes a high-order degradation modeling approach that overcomes the limitations of traditional first-order models, enabling more accurate simulation of complex, multi-stage and multi-factor coupled image degradation processes in real-world scenarios.

Through systematic analysis of interaction relationships among various degradation mechanisms, we extend the conventional first-order model to a more expressive second-order degradation model. This model achieves precise simulation of real degradation processes through the following innovative designs:

Incorporation of temporal characteristics in the degradation process to simulate multi-stage degradation in practical imaging systems;

Establishment of an adaptive coupling mechanism for degradation parameters to reflect nonlinear superposition effects of different degradation factors;

Achievement of optimal balance between computational complexity and model accuracy through a carefully designed second-order approximation scheme.

As illustrated in Figure 3, the constructed second-order degradation model generates low-resolution images that better approximate real-world scenarios through meticulously designed degradation path combinations. Training datasets synthesized based on this model effectively enhance the generalization capability of super-resolution networks under complex degradation conditions. Experimental results confirm that compared to networks trained with traditional degradation models, those trained with the second-order degradation model demonstrate significant advantages across various real-world test datasets.

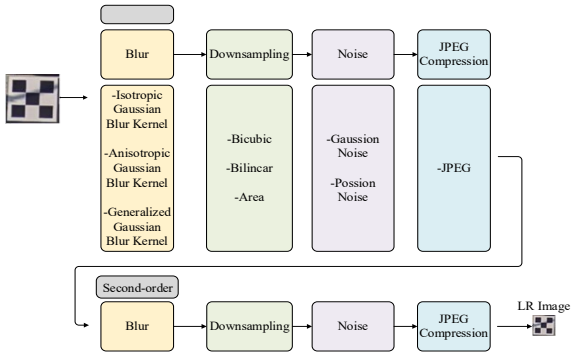


Figure 3. Advanced-order image deterioration model

2.4 *Generator Architecture*

The generator network begins with a primary feature extraction module, which takes the low-resolution (LR) image obtained from the high-order degradation model as input and extracts shallow features using a 3×3 convolutional layer. This process is formulated as:

$$F_{primary}=Conv(I_{LR}) \tag{1}$$

The advanced feature extraction module, located at the core of the generator, serves as a bridge between shallow and deep representations. It extracts multi-level high-level features through stacked fundamental blocks for subsequent reconstruction. Based on the original Real-ESRGAN network, we improve this module by introducing:Coordinate Attention Residual-in-Residual Dense Blocks (CARRDBs),a feature fusion layer and a global spatial attention block.

The CARRDB is an enhanced version of the original RRDB, incorporating a Coordinate Attention (CA) submodule after each Dense Block to refine local feature extraction. All activation functions use Leaky ReLU for faster convergence. The output of each CARRDB is fed into the next residual group for deeper feature extraction while being directly propagated to the feature fusion layer.

The feature fusion layer aggregates all CARRDB outputs through channel-wise concatenation, followed by a 1×1 convolution for dimensionality reduction:

$$F_{GF} = H_{GFF}(Concat(G_1, \dots, G_D)) \tag{2}$$

The first component of this module is an upsampling layer that performs convolutional operations on F_{high}, followed by pixel shuffling to generate high-resolution feature maps from low-resolution inputs. This process can be mathematically represented as:

$$F_{up} = Shuffle(Conv(F_{high})) \tag{3}$$

Where : $F_{primary}$ denotes the primary output features, $Conv(\cdot)$ represents the convolutional operation, I_{LR} indicates the input low-resolution image, F_{GF} corresponds to the advanced fused features with output size of, $H_{GFF}(\cdot)$ stands for the convolutional operation, $Concat(\cdot)$ signifies the feature concatenation operation, G_i refers to the output features of the i-th CARRDB (Cascaded Residual-in-Residual Dense Block),

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

F_{up} represents the upsampling output result, *Shuffle*() denotes the pixel shuffle operation.

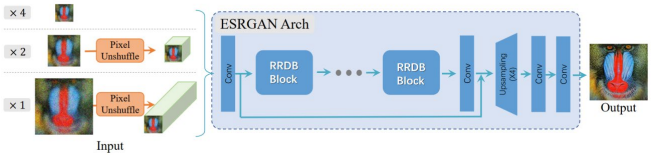


Figure 4. The architectural configuration of the conventional Real-ESRGAN generator

2.5 *Attention Mechanism*

Building upon the Real-ESRGAN framework, this study incorporates two attention mechanisms—Coordinate Attention (CA) and Spatial Attention (SA)—to enhance the network's image reconstruction performance.The implementation details of these attention mechanisms are elaborated below.

The Coordinate Attention (CA) mechanism captures long-range dependencies along two spatial directions while preserving precise positional information. This capability facilitates improved extraction of fine-grained features. Compared to alternative attention mechanisms, CA exhibits advantages such as fewer parameters and easier implementation. A schematic diagram of its structure is presented in Figure 5.

The coordinate attention mechanism is implemented through two distinct computational stages. Initially, channel attention is decomposed into dual one-dimensional feature encodings along the orthogonal X-axis and Y-axis directions, thereby addressing the inherent limitation of conventional spatial pooling methods in preserving precise positional information. For an input feature map of dimensions C×H×W, directional average pooling operations are independently performed along each spatial axis, generating orientation-aware feature representations $z_c^h(h)$ and $z_c^w(w)$, as formally expressed.

Subsequently, the coordinate attention weights are generated through the following procedure: The output features from both directional encodings undergo concatenation, followed by channel dimensionality reduction via a 1×1 convolutional layer and nonlinear activation. These operations yield the intermediate feature representation m , formally expressed as:

$$m = \delta(Conv[z^h, z^w]) \tag{4}$$

The intermediate feature m is then partitioned into two separate feature tensors, m^h and m^w, which subsequently undergo channel-wise dimensionality expansion via independent 1×1 convolutional layers. These expanded features are activated using the Sigmoid function to generate coordinate-level attention weights. This process is formally expressed as:

$$\begin{cases} g^h = \sigma\left(Conv_h(m^h)\right) \\ g^w = \sigma\left(Conv_w(m^w)\right) \end{cases} \tag{5}$$

Where : g^h and g^w denote the attention weight maps along the X-axis and Y-axis directions, respectively. By applying these weights to the input features, the calibrated output feature map is obtained as:

$$y_c(i,j) = x_c(i,j) * g_c^h(i) * g_c^w(j) \tag{6}$$

Where : $x_c(i,j)$ and $y_c(i,j)$ denote the values at coordinate (i,j) in channel c of the input and output feature maps, respectively.

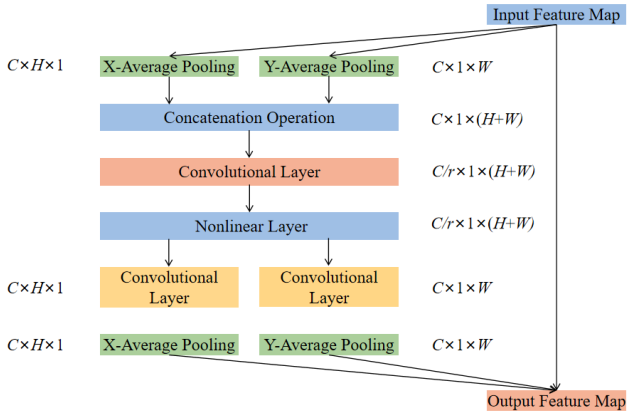


Figure 5. Architectural illustration of the Coordinate Attention module

The spatial attention mechanism captures critical spatial information within feature maps by dynamically weighting different regions of the image based on their relative importance. This enhances the model's ability to focus on semantically significant spatial locations, thereby improving the effectiveness of super-resolution reconstruction. The fundamental architecture of this mechanism is illustrated in Figure 6.

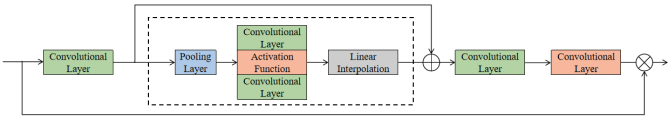


Figure 6. Architectural illustration of the Spatial Attention Module

For the input feature map X, a 1×1 convolutional layer is first applied to perform channel dimensionality reduction, yielding the output feature map X', as formally expressed by:

$$X' = Conv(X) \tag{7}$$

Subsequently, a 7×7 max-pooling operation with stride 3 is applied to X', followed by two 3×3 convolutional layers to extract spatial attention features. To restore the feature map to its original dimensions, bilinear interpolation F_{linear} is employed for upsampling. The resultant features are then element-wise summed with X', yielding the intermediate feature representation X_{mid}, as mathematically formulated below:

$$X_{mid} = F_{linear}\left(Conv\left(ReLU\left(Conv\left(F_{pool}(X')\right)\right)\right)\right) + X' \tag{8}$$

Finally, a 1×1 convolutional layer is employed to restore the channel dimensionality of the output features to match the input feature map, followed by activation via the Sigmoid function σ to generate the final spatial attention weights. These weights are then multiplied element-wise with the input feature map X to produce the output feature map Z, as mathematically formulated below:

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

$$Z = \sigma\big(F(X_{mid})\big)\odot \mathrm{X} \tag{9}$$

2.6 *Discriminator Architecture*

The high-order degradation model employed in this study exhibits significantly greater complexity than conventional low-order degradation models, inevitably leading to an orders-of-magnitude increase in computational demands. To address the consequent challenges in discriminator design, we propose an innovative U-Net architecture with spectral normalization (SN) to replace traditional VGG-based discriminators (as illustrated in Figure 7). This architectural modification offers two key advantages:

First, the spectral normalization technique effectively constrains the Lipschitz constant of network parameters, ensuring convergence within a predefined parameter space, thereby significantly enhancing training stability. Second, the U-Net's fully convolutional architecture enables pixel-wise discrimination, which maintains global semantic coherence while facilitating refined evaluation of local texture details, ultimately improving the visual realism of generated images.

Experimental results demonstrate that the proposed design achieves substantial improvements in modeling fine-grained features without compromising overall image quality.

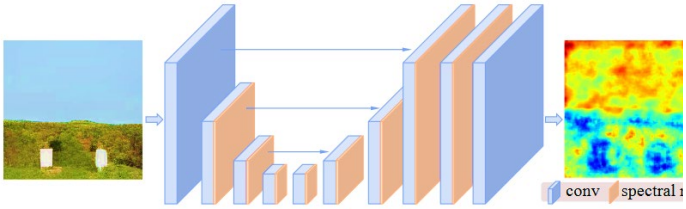


Figure 7. Spectral Normalization-incorporated U-Net Structure

3 CASE STUDIES AND ANALYSIS

3.1 *System Configuration and Dataset Composition*

To address the lack of specialized image datasets for hydraulic structure displacement monitoring, this study employs a transfer learning strategy. The model is pretrained on DIV2K, a widely adopted benchmark dataset in computer vision. DIV2K contains 800 high-quality 2K-resolution images (see Figure 8), encompassing diverse texture features and edge structures. This dataset has been proven to exhibit strong generalization performance in image super-resolution tasks, ensuring that the model learns robust generic feature representations before fine-tuning for hydraulic structure-specific scenarios.

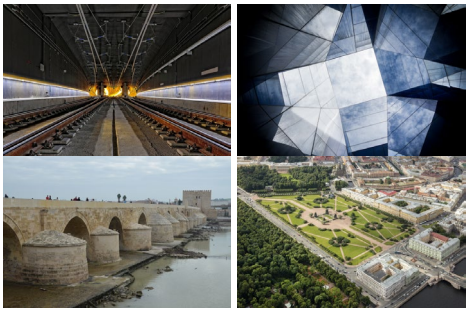


Figure 8. Representative high-resolution images from the DIV2K dataset

The network training strategy of the proposed algorithm is largely consistent with the original Real-ESRGAN model, with minor adjustments made to certain training parameters. In terms of global settings, the HR patch size is set to 256, the batch size to 48, and the Adam optimizer is employed. The network training process consists of two distinct phases. Initially, a Real-ESRNet model is trained using the L1 loss function based on a pre-trained ESRGAN model, with the iteration number set to 1×10^6 and the learning rate to 2×10^{-4}. Subsequently, the trained Real-ESRNet model is utilized as the generator initialization for the enhanced Real-ESRGAN, which is further trained with a combination of three loss functions: L1 loss, perceptual loss, and GAN loss. In this phase, the iteration number is set to 4×10^4 and the learning rate to 1×10^{-4}.

3.2 *Image Super-Resolution Reconstruction Quality Evaluation*

The proposed algorithm is trained to achieve a 4× super-resolution model, meaning it reconstructs input images at four times their original resolution. To evaluate the algorithm's performance, high-resolution (HR) images were first captured using a high-definition camera in several small- and medium-scale hydraulic engineering machine vision displacement monitoring scenarios. These images served as the reference HR images. Artificial targets within the images were cropped and selected as regions of interest (ROIs) for super-resolution reconstruction. The cropped images were then subjected to a high-order degradation process to generate low-resolution (LR) images. Subsequently, the trained model was applied to perform 4× upsampling, producing super-resolution (SR) reconstructed images.

To assess the quality of the super-resolved images, two metrics were employed: Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM). The definitions of these metrics are as follows:

Peak Signal-to-Noise Ratio (PSNR) measures the ratio between the maximum possible signal power and the mean squared error (noise power). A higher PSNR value indicates less distortion and better image quality, with the unit expressed in decibels (dB).

$$PSNR = 10lg\frac{F^2}{MSE} \tag{10}$$

Where: F denotes the maximum grayscale value of the image, MSE represents the mean squared error of the image.

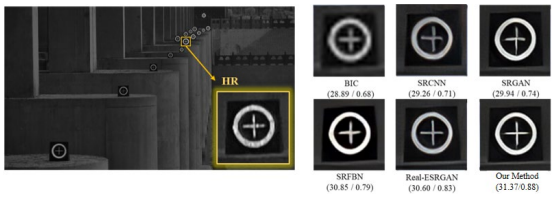
The Structural Similarity Index (SSIM) provides a comprehensive quality assessment by evaluating three key attributes: luminance, contrast, and structural fidelity. The SSIM metric is bounded within the range [0, 1], with higher values indicating better preservation of image integrity and lower distortion.

$$SSIM = l(x,y)^{\alpha}\cdot c(x,y)^{\beta}\cdot s(x,y)^{\gamma} \tag{11}$$

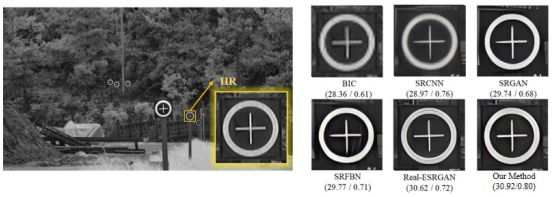
Where： $l(x,y)$ 、 $c(x,y)$ 、 $s(x,y)$ represent the luminance, contrast, and structural measure functions, respectively; α、β、γ denotes the adjustment parameter.

To validate the superiority of the proposed method, five super-resolution approaches - BIC, SRCNN, SRGAN, SRFBN, and Real-ESRGAN - were applied to the LR images for comparative performance evaluation, with the results illustrated in Figure 9.

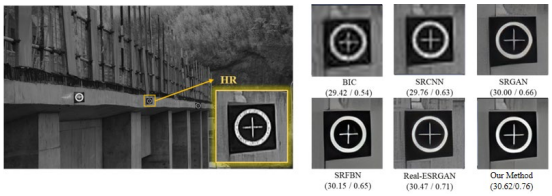
DOI: 10.3217/978-3-99161-057-1-057



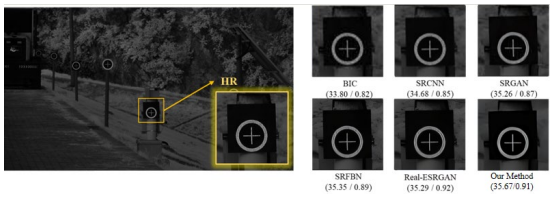
(a) Scenario 1: Machine Vision Surveillance Imagery of a Control Gate



(b) Scenario 2: Machine Vision Surveillance Imagery of a Reservoir



(c) Scenario 3: Machine Vision Monitoring Image of a Ship Lock



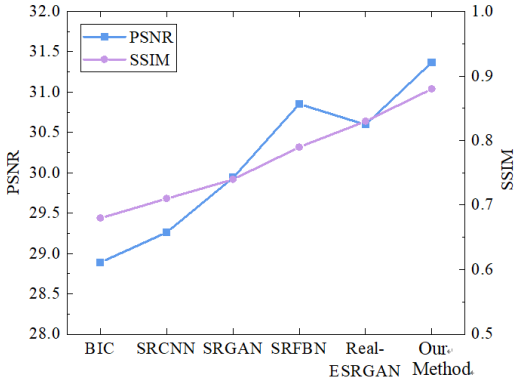
(d) Scenario 4: Machine Vision Monitoring Image of a Reservoir1

Figure 9. Super-Resolution Reconstruction Results for the Four Scenarios

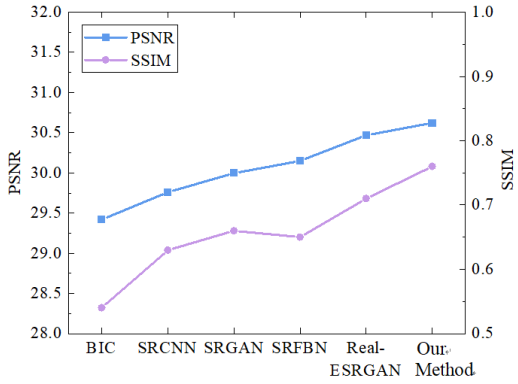
The figure provides an intuitive comparison of the image reconstruction performance among different super-resolution methods. Among them, the interpolation-based BIC method performs poorly, merely enlarging the original low-resolution image without effectively restoring the texture and edge details of the target in the image. In contrast, several deep learning-based methods yield significantly better results, though with notable variations in performance. While the four other algorithms (excluding the proposed method) can reconstruct the main contours of the image reasonably well, they still fall short in recovering fine textures and edge details.

By comparison, the proposed algorithm demonstrates the best performance. When compared to the original high-resolution (HR) image, the super-resolved (SR) image generated by our method achieves a more realistic restoration in terms of both overall visual quality and fine texture details. Moreover, owing to the improvements in feature fusion and attention mechanisms, our approach delivers more refined edge and contour reconstruction.

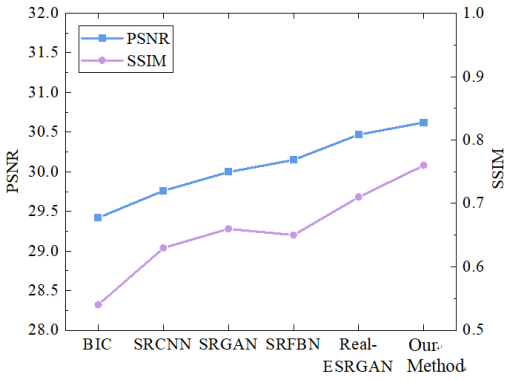
To further quantify the performance differences, the PSNR and SSIM evaluation curves of images reconstructed by different super-resolution algorithms are plotted in Figure 10. The results clearly indicate significant disparities among the algorithms, with the proposed method exhibiting superior performance in both metrics.



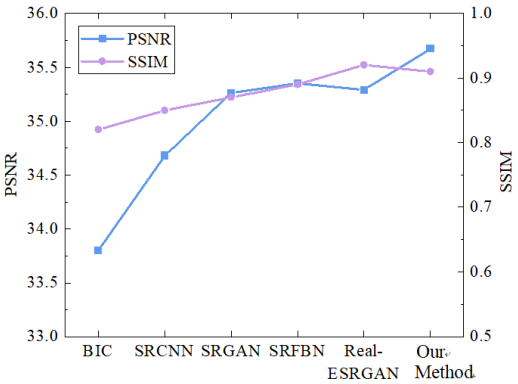
(a) Image Quality Assessment for Scenario 1



(b) Image Quality Assessment for Scenario 2



(c) Image Quality Assessment for Scenario 3



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

(d) Image Quality Assessment for Scenario 4

Figure 10. Quality validation of super-resolution reconstructed images across four distinct scenarios

3.3 *Verification of Displacement Monitoring Accuracy Improvement*

The purpose of image super-resolution reconstruction is to compensate for insufficient resolution from an algorithmic perspective when the hardware capabilities of machine vision cameras are limited or the available image resolution is low, thereby improving displacement monitoring accuracy. Therefore, it is necessary to validate the effectiveness of image super-resolution through machine vision displacement monitoring experiments.

To accurately quantify and analyze the improvement in displacement monitoring accuracy, this study selects an open-source experimental dataset for investigation. In 2018, the University of California, San Diego conducted a series of shake table tests to study the lateral response characteristics of cold-formed steel frame building structures. The related design schemes, test reports, videos, and data were all publicly released. This experiment provides comprehensive video recordings and measured data from various sensors, making it highly suitable for verifying displacement monitoring accuracy in this study.

For this research, a specific shear wall specimen from the test series was selected, and its dynamic test video data under seismic loading were analyzed. The vibration input was scaled from the recorded ground motion of the 1994 Northridge earthquake in Los Angeles, USA, ensuring the specimen remained elastic throughout the test. The experimental setup of the shake table test is illustrated in Figure 11. Specifically:

Displacement sensors were installed on the side of the load-transfer beam at the top of the specimen to measure lateral displacements induced by horizontal loading.

A machine vision camera** was positioned directly in front of the specimen to record the entire dynamic testing process.

This configuration allows for a comparative evaluation of displacement measurements obtained from traditional sensors and machine vision-based methods, facilitating an assessment of the accuracy enhancement achieved through super-resolution reconstruction.

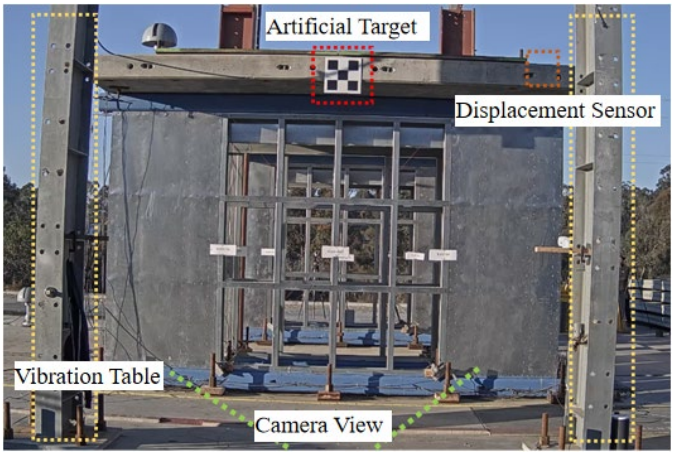


Figure 11. Overview of the Shaking Table Field Test

The checkerboard-pattern artificial target installed on the load-transfer beam at the top of the specimen served as the region of interest (ROI) for machine vision-based displacement monitoring. The specific geometry and dimensions of this target are illustrated in Figure 12.The sensor-measured displacement data recorded during the vibration test are presented in Figure 13.

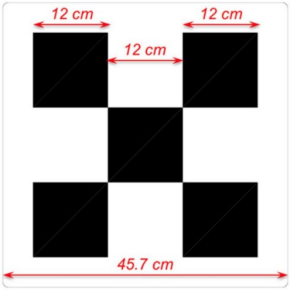
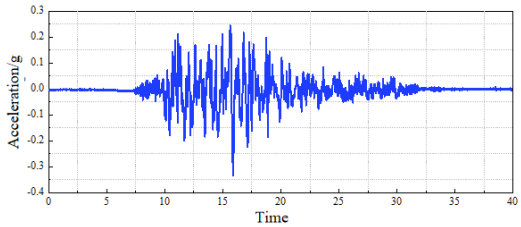
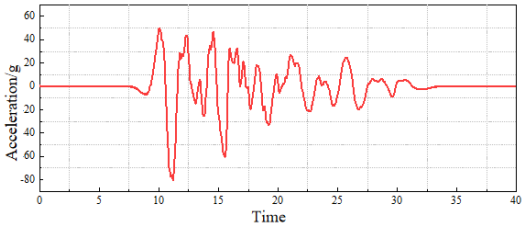


Figure 12. Shape and Dimensions of the Artificial Target



(a) Accelerometer measurement data



(b) Displacement sensor measurements

Figure 13. Experimental sensor data recorded during vibration testing

During the experimental testing process, the images were captured at close range using a high-quality camera, resulting in high-resolution footage that meets the requirements of machine vision-based displacement monitoring under normal conditions. Therefore, the actual captured monitoring images were treated as high-resolution (HR) images, and displacement monitoring was performed on the HR image sequence. The tracking target was a checkerboard-pattern artificial target, and the calculated results are presented in the following figures and tables. The results demonstrate that the machine vision-based displacement monitoring using high-resolution images achieved high accuracy, closely matching the measured displacement values from sensors, thereby validating the effectiveness of the machine vision approach.

To evaluate the accuracy improvement effects of different super-resolution methods, the original HR images (120×120 pixels) of the target region were first degraded according to the higher-order degradation model proposed in this study, generating corresponding low-resolution (LR) images (30×30 pixels) as the baseline LR images for the accuracy enhancement validation experiment (all super-resolution algorithms were

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

applied to these LR images for enhancement). Subsequently, the bicubic interpolation (BIC) method was employed to perform 4× upsampling super-resolution reconstruction on the LR images, producing corresponding BIC super-resolution (SR) images (120×120 pixels) for comparative analysis with other super-resolution algorithms. The entire process is illustrated in Figure 14 below.

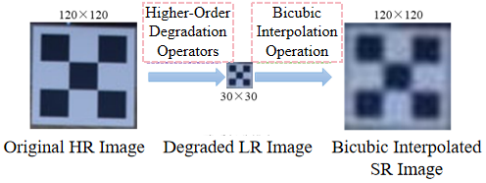


Figure 14. Processing results of artificial target images

For the low-resolution (LR) images, the proposed algorithm in this study was employed to perform super-resolution (SR) reconstruction, and the results were compared with those of other SR methods, as illustrated in the following figure. As shown in Figure 15, the visual differences among the outputs of different methods can be easily observed. Among them, the BIC and SRCNN super-resolution algorithms exhibited the poorest performance, with noticeable blurring visible to the naked eye. The other four algorithms achieved better reconstruction results, but the proposed method outperformed them all, achieving the highest scores in both PSNR and SSIM metrics. When comparing the super-resolved images generated by the proposed method with those obtained from the original Real-ESRGAN (before improvements), it is evident that the proposed method produces images with sharper edges and more detailed textures, demonstrating the effectiveness of the introduced feature fusion and attention mechanism enhancements.

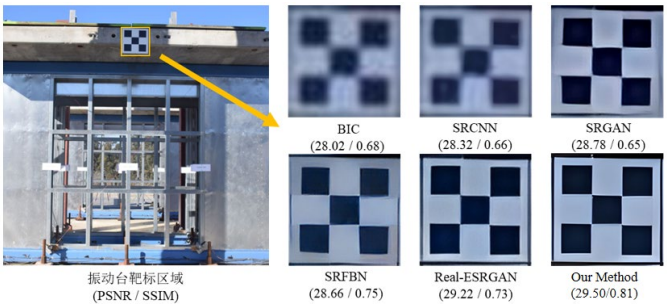


Figure 15. Super-resolution reconstruction results of the target region

To further analyze the quality improvement effects of image super-resolution, corner detection was performed on different types of images using OpenCV's goodFeaturesToTrack detector, followed by cornerSubPix refinement for sub-pixel accuracy. All detection parameters remained consistent throughout the process, and the resulting corner detection outcomes are illustrated in Figure 16. Given the noticeable differences in performance among the algorithms, the superiority of each method can be clearly determined based on the number and positional accuracy of detected corners.

Taking the original high-resolution (HR) image as a reference, a total of 16 inner checkerboard corners and 5 outer corners were detected, with their positions precisely aligned to the edges and intersections of the checkerboard pattern. The comparison among different super-resolution algorithms revealed significant discrepancies. For instance, BIC and SRCNN exhibited evident corner misalignment and missing detections, failing to fully capture the expected number of corners. In contrast, the proposed method demonstrated the best performance, detecting the same number of corners as the HR image while maintaining accurate positional correspondence, further validating its effectiveness.

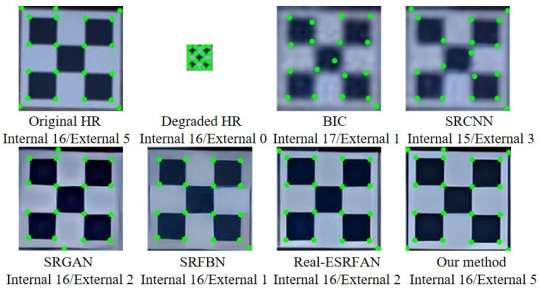
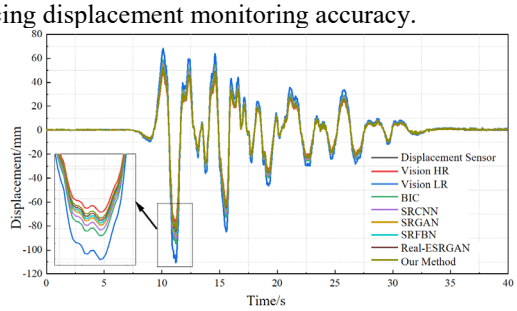


Figure 16. Corner detection results for images reconstructed using different super-resolution methods

Subsequently, target tracking and displacement calculation were performed on the reconstructed image sequences obtained from the aforementioned super-resolution methods, yielding the corresponding displacement time-history curves for each method, as illustrated in Figure 17. The results demonstrate that the low-resolution (LR) images without any processing exhibited degraded tracking performance due to substantial loss of effective information and noise contamination, leading to unsatisfactory displacement monitoring results. Specifically, the displacement curve of the LR images displayed significant fluctuations and drift phenomena, particularly in the intermediate segment with larger vibration amplitudes (as clearly observed in the zoomed-in subplot). The resulting errors substantially exceeded the acceptable threshold, rendering the displacement monitoring results unreliable.

In contrast, the six curves derived from super-resolution-processed images exhibited significantly improved performance, mitigating the accuracy degradation caused by image resolution reduction to varying degrees. Among these, the Bicubic Interpolation (BIC) algorithm demonstrated relatively inferior performance, while the proposed feature-fusion-enhanced Real-ESRGAN algorithm with attention mechanisms achieved the best results. The displacement curve generated by the proposed method closely aligned with both the sensor-measured displacement curve and the vision-based measurement curve obtained from high-resolution (HR) images, outperforming even the original Real-ESRGAN algorithm. This comparative analysis further validates that the proposed model improvements contribute effectively to enhancing displacement monitoring accuracy.



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

Figure 17. Vibration displacement monitoring results using different super-resolution methods

Finally, four evaluation metrics—Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Root Mean Square Error (RMSE), and the Coefficient of Determination (R^2)—were employed for further quantitative assessment. The corresponding residual box plots and evaluation metric radar charts are presented in Table 1 and Figure 18, respectively. The results demonstrate that the proposed method consistently outperformed all other approaches across all evaluation metrics, achieving an R^2 value of 0.9975, which is remarkably close to the monitoring results obtained from the high-resolution (HR) images. Although the proposed method cannot fully compensate for the degradation caused by image downsampling, it significantly mitigates the impact of resolution reduction compared to other methods, exhibiting superior performance.

The residual box plots reveal that the monitoring results based on HR images exhibited the highest accuracy, with residuals fluctuating only within a very narrow range. Considering that this case involves large-amplitude vibration displacement, which inherently introduces error drift in target tracking, the BIC super-resolution algorithm displayed the largest residual fluctuations, while the residuals of other algorithms also exceeded 5 mm. In contrast, the proposed method effectively confined the residuals within 5 mm, demonstrating a substantial improvement in monitoring accuracy over competing approaches.

Table 1 Displacement monitoring accuracy evaluation across different super-resolution methods

Test Category	MAE	MAPE	RMSE	R^2
Visual HR	0.1016	0.3165	0.2330	0.9998
Visual LR	3.8222	0.4959	6.6977	0.9182
BIC	1.9549	0.4165	3.3559	0.9721
SRCNN	1.4882	0.3930	2.5221	0.9828
SRGAN	1.1149	0.3730	1.8568	0.9910
SRFBN	0.9283	0.3626	1.5254	0.9930
Real-ESRGAN	0.7418	0.3519	1.1959	0.9955
Proposed Method	0.5552	0.3409	0.8701	0.9975

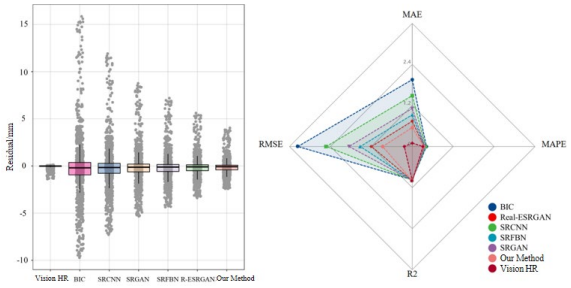


Figure 18. Error evaluation visualization across different super-resolution methods

4 CONCLUSION

Departing from conventional approaches that rely on hardware upgrades (e.g., higher-resolution cameras, precision optical lenses) or subpixel algorithm refinement, this research innovatively explores image super-resolution (SR) reconstruction based on deep learning, proposing a software-algorithm-level solution for breakthrough accuracy enhancement in displacement monitoring. The key research contributions and findings are systematically summarized as follows:

A thorough analysis of image degradation processes and super-resolution reconstruction principles was conducted. An SR reconstruction model was innovatively developed using a generative adversarial network (GAN) architecture. By establishing a multi-factor coupled higher-order degradation model, the study accurately simulates complex imaging processes in real-world engineering environments, significantly improving the algorithm's adaptability to various degraded images.

The study innovatively integrates Coordinate Attention (CA) and Spatial Attention (SA) mechanisms into the Real-ESRGAN framework, constructing a feature fusion attention network. The CA mechanism enables channel-space feature co-optimization, while the SA mechanism enhances the representation of critical regions. Without modifying hardware configurations, this approach significantly improves target feature discriminability, providing a reliable foundation for high-precision displacement monitoring.

The proposed method was rigorously validated through engineering case studies, assessing both image quality metrics and displacement monitoring accuracy. Comparative experiments with five state-of-the-art algorithms demonstrate the superior performance of the proposed method, which effectively restores edge and contour details in reconstructed images. Moreover, the displacement monitoring results derived from the SR-enhanced image sequences exhibit significantly improved accuracy, offering an innovative technical pathway to overcome the precision limitations in machine vision-based monitoring of hydraulic structures.

REFERENCES

[1] Zhao, Y., Wang, H., Ma, H., et al. (2022). Construction concept of China's "Double T" water network economic pattern. Journal of Hydraulic Engineering, 53 (11), 1271-1279, 1290.

[2] Wang, J., Zhao, H., & Ye, Y. (2018). Smart water network engineering: The engine driving China's water governance modernization. Journal of Hydraulic Engineering, 49(9), 1148–1157.

[3] YANG L, SU H, WEN Z. Improved PLS and PSO methods-based back analysis for elastic modulus of dam[J]. Advances in Engineering Software, 2019, 131: 205–216.

[4] Ban, Y., Liu, Y., Wang, N., et al. (2024). A review of measurement methods for the main reflector surface of radio telescope antennas. Scientia Sinica: Physica, Mechanica & Astronomica, 54(1), 23–37.

[5] WANG X, ZHAO Q, XI R, et al. Review of bridge structural health monitoring based on GNSS: from displacement monitoring to dynamic characteristic identification[J]. IEEE Access, 2021, 9: 80043–80065.

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

[6] Liu, Y. (2023). Research on construction monitoring technology for the swivel bridge of the overpass crossing Yanshi Railway [Master's thesis, North China University of Technology].
[7] Li, G. (2023). Thoroughly implement the spirit of the 20th CPC National Congress and solidly promote high-quality development of water conservancy in the new stage: Speech at the 2023 National Water Conservancy Work Conference. China Water Power & Electrification, 2023(2), 1–11.
[8] Liu Z.P., Fu H., Guo X.L., et al. 2017. Integrated dual-frequency radar measurement system for ice-water conditions. Journal of Hydraulic Engineering 48(11): 1341-1347.
[9] ZHOU Q, LI Q-S, HAN X-L, et al. Improvement of GPS displacement measurement accuracy for high-rise buildings by machine learning[J]. Journal of Building Engineering, 2023, 78: 107581.
[10] NURKOWSKI J, NOWAKOWSKI A. Inductive sensor for measuring linear displacement and velocity – Version with stationary magnetic core[J]. Measurement, 2023, 222: 113675.
[11] HAN J, XIONG G, LIU J. Detection and analysis of pavement-section based on laser displacement sensor[J]. Sensors, 2023, 23(15): 6758.
[12] SUH Y S. Laser sensors for displacement, distance and position[J]. Sensors, 2019, 19(8): 1924.
[13] LIU T, LEI Y, MAO Y. Computer vision-based structural displacement monitoring and modal identification with subpixel localization refinement[J]. T.-C. Huynh. Advances in Civil Engineering, 2022, 2022: 1–11.
[14] LUO L, FENG M Q, WU Z Y. Robust vision sensor for multi-point displacement monitoring of bridges in the field[J]. Engineering Structures, 2018, 163: 255–266.
[15] KARSZNIA K, OSADA E, MUSZYŃSKI Z. Real-time adjustment and spatial data integration algorithms combining total station and GNSS surveys with an earth gravity model[J]. Applied Sciences, 2023, 13(16): 9380.
[16] ZHOU J, HE L, LUO H. Real-time positioning method for uavs in complex structural health monitoring scenarios[J]. Drones, 2023, 7(3): 212.
[17] LI X, ZHONG B, LI J, et al. Inversion of GNSS vertical displacements for terrestrial water storage changes using slepian basis functions[J]. Earth and Space Science, 2023, 10(2): e2022EA002608.
[18] CAO S, LU X, SHEN S. GVINS: tightly coupled gnss–visual–inertial fusion for smooth and consistent state estimation[J]. IEEE Transactions on Robotics, 2022, 38(4): 2004–2021.
[19] KIM K-H, JUNG H-K. Development of a remote displacement measuring laser system for bridge inspection[J]. Sensors, 2022, 22(5): 1963.
[20] PAZIEWSKI J, STEPNIAK K, SIERADZKI R, et al. Dynamic displacement monitoring by integrating high-rate GNSS and accelerometer: on the possibility of downsampling GNSS data at reference stations[J]. GPS Solutions, 2023, 27(3): 157.
[21] FA G, LI K, CAO T. Improving vibration monitoring of structures using theodolites with built-in image sensors[J]. International Journal of Structural Stability and Dynamics, 2023, 23(11): 2350125.
[22] KHUC T, CATBAS F N. Completely contactless structural health monitoring of real-life structures using cameras and computer vision: structural health monitoring using computer vision[J]. Structural Control and Health Monitoring, 2017, 24(1): e1852.
[23] YOON H, ELANWAR H, CHOI H, ET al. Target-free approach for vision-based structural system identification using consumer-grade cameras: Target-Free Vision-based structural system identification[J]. Structural Control and Health Monitoring, 2016, 23(12): 1405–1416.
[24] DONG C Z, YE X W, JIN T. Identification of structural dynamic characteristics based on machine vision technology[J]. Measurement, 2018, 126: 405–416.
[25] BROWNJOHN J M W, XU Y, HESTER D. Vision-Based bridege deformation monitoring[J]. Frontiers in Built Environment, 2017(3): 23.
[26] BOCIAN M, NIKITAS N, KALYBEK M. Dynamic performance verification of the Rędziński Bridge using portable camera-based vibration monitoring systems.[J]. Archives of Civil and Mechanical Engineering, 2023(23): 1–19.
[27] KOHUT P, HOLAK K, MARTOWICz A, et al. Experimental assessment of rectification algorithm in vision-based deflection measurement system[J]. Nondestructive Testing and Evaluation, 2017, 32(2): 200–226.
[28] WU T, TANG L, SHAO S, et al. Cost-effective, vision-based multi-target tracking approach for structural health monitoring[J]. Measurement Science and Technology, 2021, 32(12): 125116.
[29] MOLINA-VIEDMA A J, FELIPE-SESÉ L, LÓPEZ-ALBA E, et al. High frequency mode shapes characterisation using digital image correlation and phase-based motion magnification[J]. Mechanical Systems and Signal Processing, 2018, 102: 245–261.
[30] Wang J., Huang Y., Deng Y., et al. 2021. Study on fracture characteristics of Yellow River ice based on digital image correlation method. Journal of Hydraulic Engineering 52(9): 1036-1046.
[31] GUO J, XIANG Y, FUJITA K, et al. Vision-Based Building Seismic Displacement Measurement by Stratification of Projective Rectification Using Lines[J]. Sensors, 2020, 20(20): 5775.