Enhancing Dynamic Analysis of Mechanical Systems using Dynamic Gaussian Splatting and Motion Magnification

W. Bittner [1,2], T. Slimak[1], H. Jung[2]
[1] TU Munich, TUM School of Engineering and Design,
Boltzmannstraße 15, 85748 Garching bei München, Deutschland
[2] TU Munich, TUM School of Computation, Information and Technology (CIT)
Boltzmannstraße 3, 85748 Garching bei München, Deutschland
e-mail: **william.bittner@tum.de**

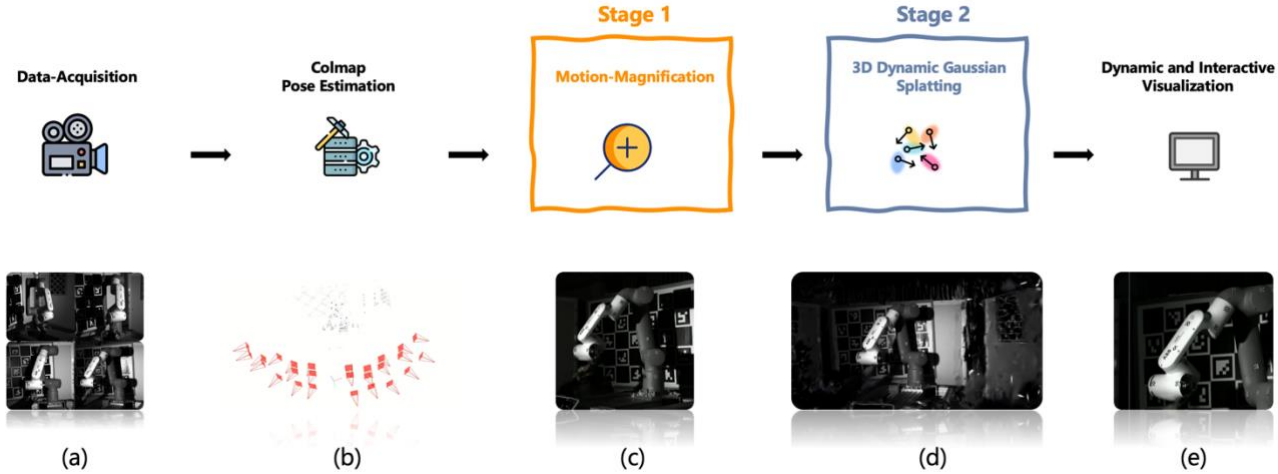

Figure 1 Overview of the overall Pipeline.
(a) Input: Multiple videos from different views of a system with subtle motion.
(b) COLMAP Pose Estimation: It provides Structure from Motion (SfM) Points and Camera Pose Estimates.
(c) Stage 1: Motion Magnification is applied independently on each input video.
(d) Stage 2: 3D Dynamic Gaussian Splatting builds a dynamic 3D representation from the motion magnified videos.
(e) Output: Dynamic 3D Gaussian representation of captured scene with magnified motion renders novel views in real-time.

Abstract:

This research combines Motion Magnification and Dynamic Gaussian Splatting to visualize 3D operational deflection shapes in mechanical systems using only video input. Motion Magnification amplifies subtle motions in video recordings, while Dynamic Gaussian Splatting enables 3D scene representation with time-dependent parameters. By integrating these techniques, the method creates a comprehensive 3D visualization of amplified structural motion, offering a contactless, full-field approach to structural dynamics analysis. The methodology was validated using both synthetic data (3D Speaker Model and robot arm) and real-world data (impact hammer-excited robotic arm). Videos from multiple viewpoints were processed using learning-based motion magnification and then used to build a 3D dynamic Gaussian model, enabling visual analysis of the system's modal parameters.

1 Introduction

Visualizing and understanding structural dynamics is crucial for optimizing designs and diagnosing structural issues in mechanical systems such as loose components or cracks. While traditional Experimental Modal Analysis (EMA) provides quantitative data on modal parameters using accelerometer measurements at discrete points, it can be challenging to intuitively visualize the full-field structural response. Our proposed method aims to complement existing techniques by providing a contactless, full-field approach to visualizing operational deflection shapes. It integrates Motion Magnification (MM) [1] with Dynamic Gaussian Splatting (DGS) [2] to magnify and visualize subtle, imperceptible motions in 3D. MM amplifies subtle motions in video data at continuous locations, overcoming the limitations of sparse accelerometer data. DGS enables the visualization of these full-field measurements in 3D space, providing a visual intuitive and interactive representation of the system's operational deflection shapes. This integration offers a powerful tool for engineers to gain deeper insights into structural dynamics.

1.1 EMA & ODS

Experimental Modal Analysis (EMA) is a well-established technique for determining the modal parameters (natural frequencies, damping ratios, and mode shapes) of a mechanical system. It involves a three-stage process: defining test conditions and exciting the structure, measuring responses using accelerometers, and analyzing the data to extract modal parameters[3][4][5]. While EMA provides precise quantitative data on modal parameters, the optimal placement of sensors is crucial for accurate results, it relies on discrete measurement points and can be time-consuming to set up. It is worth noting, that the proposed method does not aim to replace EMA.

A related but distinct approach is Operating Deflection Shapes (ODS) analysis. ODS is a way to visualize how a structure or machine moves under operational conditions. Unlike EMA, ODS tests have no measured artificial forces and only measure response vibration signals from accelerometer sensors that exhibit the super-positioned response of all excited mode-shapes. It provides structural deflection shapes which can improve the understanding of a structure's dynamic behavior during operation [6].

In contrast to current ODS Analysis, which operates on discrete accelerometer data, our proposed method provides contactless 3D ODS from full field measurements (camera recordings) of structural dynamics. By combining motion magnification techniques with 3D reconstruction, we enhance the visualization of operational deflection shapes, providing intuitive insights into the overall motion of the structure without the need for physically attached sensors and the required expert knowledge where to place the limited number of sensors on the dynamic structure.

It's important to note that while the proposed method provides an enhanced visualization of structural dynamics, it does not directly determine modal parameters such as natural frequencies, mode shapes, or damping ratios. The output is a visual representation of operational deflection shapes (ODS), which can provide intuitive insights into structural behavior but should not be considered equivalent to the quantitative results obtained from traditional EMA techniques. It can rather complement EMA by providing indication on good placements of the sensors for EMA or give insights into a dynamic system when an ODS analysis is sufficient. Future work will focus on extracting quantified motion data to also enable the determination of modal parameters.

1.2 Learning-based Video Motion Magnification (MM)

Learning-based MM amplifies subtle motions in videos, revealing dynamics invisible to the naked eye with a magnification factor of up to 100x [1]. A deep convolutional neural network (DCNN) [7] is trained on a large dataset of synthetically generated motion-magnified video pairs (200,000 background images with 7,000 moving foreground objects) to magnify motion changes between frames. The network learns to extract a representation of motion from video frames that can be directly manipulated to amplify subtle movements. The architecture comprises of an Encoder, that extracts shape and texture representations from each frame,

a Manipulator, that magnifies motion by amplifying the difference between shape representations of consecutive frames and a Decoder, that reconstructs the magnified frame using the manipulated shape and texture information.

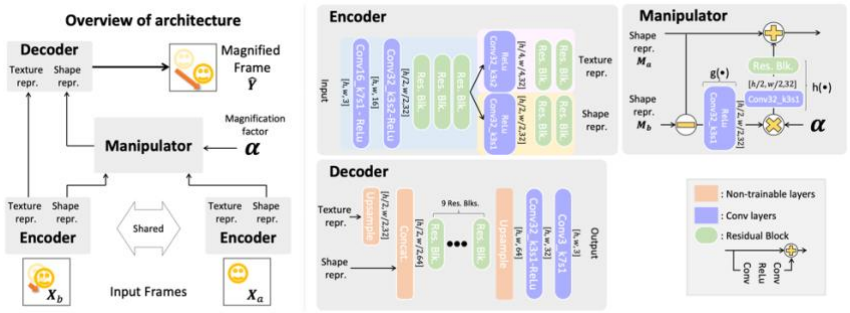

Figure 2 Network architecture of Learning-based Video Motion Magnification

Encoder G_e (·): The fully-convolutional encoder acts as a spatial filter, extracting shape $M = G_{e,\text{shape}}(X)$ and texture $V = G_{e,\text{texture}}(X)$ representations from a frame. This dual representation helps combat noise and unwanted intensity magnification and is enforced through a regularization loss during training.

Manipulator G_m(·): The manipulator takes the shape representations of two frames X_a and X_b and magnifies the motion by multiplying the shape difference $M_a - M_b$ with a magnification factor α. Some additional non-linearities $h(\cdot)$ (3×3 convolution + 3×3 residual block) and $g(\cdot)$ (3×3 convolution + ReLU [8]) help improve the quality.

$$G_m(M_a, M_b, \alpha) = M_a + h(\alpha \cdot g(M_b - M_a)) \tag{1}$$

Despite being trained only with two-frame inputs, the network shows that the shape representation is sufficiently linear with respect to displacement, making it compatible for linear temporal filters $T(\cdot)$. These can be applied pixel-wise across the temporal axis within the representation manipulator $G_m(\cdot)$, enhancing temporal coherence in magnified frames.

$$G_{m,temporal}(M(t), \alpha) = M(t) + \alpha T(M(T)) \tag{2}$$

Decoder G_d(·): The fully-convolutional decoder takes the temporally-filtered shape representation as well as the texture representation of one input frame and reconstructs the motion magnified output frame.

For more detailed information we refer the reader to the original paper by Oh. et al. [1].

1.3 (Dynamic) Gaussian Splatting (GS)

Gaussian Splatting (GS) achieves high-quality novel-view synthesis of scenes captured with multiple photos by representing scenes as collections of 3D Gaussians [9]. Each Gaussian encodes position, rotation, scale, opacity, and color. In the beginning of the optimization loop a set of 3D Gaussians is initialized with SfM Points. An interleaved optimization/density control method is used to update the overall number of 3D Gaussians and their parameters (position, rotation, …) in a scene over the course of 30k iterations. Efficient rendering is achieved through a visibility-aware tile-based algorithm with fast GPU sorting that "splats" the 3D Gaussians from 3D space to the 2D camera view plane.

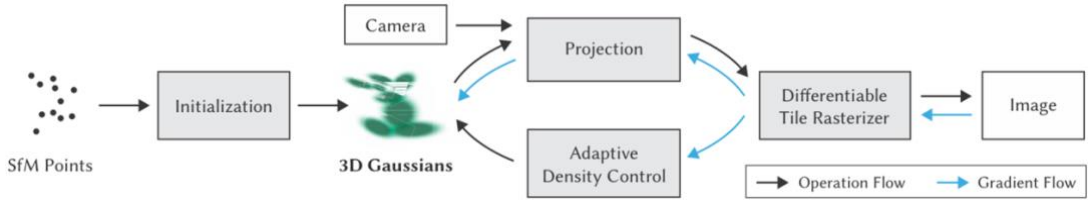

Figure 3 Method-Overview of 3D Gaussian Splatting

Dynamic Gaussian Splatting (DGS) [2] extends GS to dynamic scenes by parametrizing the 3D Gaussian parameters; 3D center and 3D rotation on a per-timestep basis over the input time-sequence, while the other parameters; 3D scale, opacity and color are fixed after the first timestep. This enables the representation and rendering of moving objects and scenes given video-sequences from multiple viewing angles. As in the case of (static) Gaussian Splatting, the 3D Gaussians are initialized with SfM-Points and optimized for 15k iterations in the first timestep. In every subsequent timestep the 3D Gaussians are initialized from the last timestep and optimized for 2k iterations to model the movement. Additional Loss-Functions (local rigidity, rotation, long-term local isometry) are introduced to regularize the movement of the 3D Gaussians between consecutive frames.

```
Initialize GaussianModel with COLMAP SfM-Points
for each timestep t do
    if t = 1 then
        Iterations ← 15000
    else
        Iterations ← 2000
    end if
    for i = 1 to Iterations do
        Render Scene using Gaussian Model
        Compute Loss between Rendered Scene and Actual Frame at timestep t
        Backpropagate Loss to Update Gaussian Parameters:
            • Centroid
            • Rotation
            • Scale (only for first timestep)
            • Opacity (only for first timestep)
            • Color (only for first timestep)
        Update Gaussian Model with new parameters
    end for
    Initialize Gaussian Model for next timestep with Parameters from last timestep
end for
```

Figure 4 Algorithm of Dynamic 3D Gaussians Training Routine

For more detailed information we refer the reader to the original paper by Luiten. et al. [2].

1.4 Hybrid image- and accelerometer-based Mode Shape magnification

To contextualize our work within the field and highlight the unique contributions of our approach it helps to briefly mention Hybrid image- and accelerometer-based Mode-shape magnification [10], that focuses on a hybrid approach that enables the identification of structural dynamics below the sub-pixel range and beneath the image noise floor. This method combines experimental modal analysis (EMA) with a gradient-based optical flow technique to capture the dynamic response of structures under excitation. The technique employs a triangle mesh to subdivide the image of the vibrating structure, which is then warped according to the identified motion. By integrating data from conventional vibration sensors (e.g., accelerometers) with image processing, this approach achieves impressive magnification factors of up to 40,000 times.

While this hybrid method provides valuable insights into structural dynamics, it relies on a predefined 3D mesh of the structure, which is not always available upfront and can be time-consuming to create. Our proposed method, in contrast, requires no prior 3D model and is well-suited for analyzing such complex structures. By leveraging multiple camera views and Dynamic Gaussian Splatting, we reconstruct the 3D motion directly from the video data, enabling an intuitive visualization of operational deflection shapes without the need for a predefined mesh or direct quantitative modal analysis.

2 Method

Our proposed 2-stage method (Motion Magnification and Dynamic 3D Gaussian Splatting) in **Figure 1** achieves a motion-magnified, interactive 3D visualization of the operational deflection shapes of mechanical systems from camera captures of a dynamic structure from multiple viewpoints. The following section outlines the steps applied in this work to generate the presented results.

a) **Synthetic Data Generation:** Synthetic datasets were generated by modeling, lighting, and animating 3D scenes with dynamic objects within Blender [11]. These scenes were then rendered

from multiple viewpoints for a predefined number of timesteps, simulating multi-view video capture.

Real-World Data Acquisition: Real-world data was captured using a single high-speed camera (Photron Fastcam Nova S6 800k) at 640 fps. To determine the optimal sampling rate, 6400 fps were recorded, with subsampling applied to empirically test the performance of different step sizes. Too large steps result in jagged motion, while too small steps result in too little frame-to-frame motion. A robotic arm served as the test object, illuminated by a strong LED to ensure adequate exposure during the short capture times. Excitation of the robotic arm was induced via an impact hammer at a designated location. To obtain multi-view data, the excitation experiment was repeated for each of the 28 camera viewpoints, with subsequent synchronization of the video sequences in post-processing. While this sequential acquisition process required approximately 3 hours, a multi-camera setup could significantly reduce this time. The impact excitation was maintained consistently throughout all experiments. The high frame rate and uncompressed nature of the video data resulted in large file sizes (~10GB per recorded second), leading to extended data saving times (~2 minutes per recording). This bottleneck in the acquisition process could be mitigated by employing a multi-camera setup for simultaneous capture or by reducing the data volume through lower frame rates or compression.

For later 3D reconstruction at Stage 2, good camera pose estimates are necessary, which are obtained using COLMAP **(Colmap Pose Estimation)**. COLMAP is a photogrammetric technique that estimates the extrinsic *w2c* (rotation *R* and translation *t*) and intrinsic *k* (focal length, principal point, skew) parameters of the camera for each image and reconstructs 3D structures with sparse Structure from Motion (SfM) points from a series of 2D images taken from different viewpoints by identifying and matching feature points across the images. The set of SfM points obtained serves as the initialization of the 3D Gaussians at the beginning of Stage 2 [12], [13].

$$\mathbf{w2c} = \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0}^{\mathrm{T}} & 1 \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} & x \\ r_{21} & r_{22} & r_{23} & y \\ r_{31} & r_{32} & r_{33} & z \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{3}$$

In **Stage 1 (Motion Magnification)** a pre-trained deep convolutional neural network (DCNN) amplifies subtle motions in each video sequence individually based on the work of Oh et al. [1] by factor 100x in static mode. For higher RGB-noise levels of the input image sequence temporal filtering is applied to smooth out the noise. Furthermore, the original code was refactored to be compatible with TensorFlow 2 for accelerated computing of multiple parallel streams.

In **Stage 2 (Dynamic 3D Gaussian Splatting)** the motion-magnified image sequences and additional metadata (initial point cloud and camera poses) are used to optimize the Dynamic 3D Gaussian scene. To improve the quality of the Gaussian Model, the Dynamic Gaussians are conditioned on many photos of the static scene for the first timestep before training the scene dynamics with sparse videos for the subsequent timesteps. Training the Dynamic 3D Gaussian Model is computationally the most expensive step and depends on the temporal length of the dynamic scene, but not the number of training views. Training a Dynamic 3D Gaussian Model for 50 timesteps (~2s of motion) takes 45mins on a RTX4090. Any additional timestep takes about 45 seconds to train. To make the synthetic dataset compatible with the DGS code, it needed to be adapted. Specifically, the raw input videos must be accompanied by a meta.json file that provides information about the camera poses and the respective file-paths of the input data.

b) Finally, the learned Dynamic 3D Gaussian representation of the captured scene can interactively visualize the magnified object-motion in real-time **(Dynamic and Interactive Visualization)**, thus providing an intuitive understanding of the operating deflection shapes. In this context, “interactively” means, that the user can freely move and rotate the camera to render novel views of the captured Dynamic 3D Gaussian scene in real time.

3 Datasets and Experimental Parameters

To evaluate our method, three datasets were generated/acquired.

Synthetic Dataset 1 involved a simple 3D speaker model with subtle membrane oscillation to quickly make first tests to validate the potential of the method and assess, whether the order of the two stages has a significant effect on the output quality

Synthetic Dataset 2 involved a 3D robot model in a realistic environment with subtle rotatory oscillation of the arm to perform a sensitivity analysis with respect to rgb noise, number of camera views and camera pose accuracy.

- For the ***number of training views sensitivity analysis,*** the number of training views available at stage 2 was altered from 3 to 50 to determine the minimum amount of camera views / visual coverage necessary for good results. Additionally, conditioning the first timestep of the DGS model (stage 2) with an additional set of camera views was tested to see whether this reduces the number of required camera views for the dynamic scene. The idea is that the initializing camera views capture the geometry of the scene while the camera views for the dynamic scene capture the dynamics. It should be noted that capturing a set of initializing camera views of the static scene to condition the DGS model (stage 2) at timestep 0 is easy to acquire in a real-world scenario which simply requires walking around the static object and recording it from different viewing angles, whereas obtaining a set of camera views for the dynamic scene either requires an expensive multi-camera setup or repeating the experiment multiple times and recording it each time from a different angle and later synchronize the asynchronously acquired videos in post.
- For the ***camera pose accuracy sensitivity analysis,*** the camera pose provided to the DGS model (stage 2) was perturbed by some Gaussian Noise of up to 20mm in translation error and 2 deg in rotation error to measure the robustness against camera pose estimation errors. These experiments aimed to assess the robustness of the pipeline to inaccuracies in camera pose estimation, a common challenge in real-world scenarios, when dealing with COLMAP camera pose estimates. Errors in camera poses directly cause inaccuracies in the 3D Gaussian model, leading to distortions, blurring and noise in the motion visualization.
- For the ***rgb noise sensitivity analysis,*** the input images to the first stage (Motion-Magnification) were perturbed by some rgb noise to impair the output of the first stage and training the 3D Gaussian Model on this perturbed data to see, how robust the method responds to rgb-noise. To add noise to each pixel (rgb-tupel in [0,1]), we introduce Gaussian noise $N\ (0,\ \sigma)$ that is added to each RGB value. Here, $\mu = 0$ is the mean, and σ is the noise level, which determines the standard deviation of the Gaussian noise. In our experiments we evaluated the metrics for rgb noise levels from 0.01 to 0.05. We trained the second stage (Dynamic 3D Gaussian Splatting) with temporally filtered motion magnified image sequences and with regular motion magnified image sequences. For further details to the temporal bandpass filter refer to [11].

The **Real-World Dataset** involved a robotic arm that was manually excited by an impact hammer to validate the method on a real-world dataset. The scene was recorded by a monochromatic Photron Fastcam Nova S6 800k high-speed camera for 1s at 640 fps from 27 viewpoints. The focal length was 50mm and the aperture closed at f/16 to avoid blurry out-of-focus regions. A strong LED was placed in front of the robotic arm to provide enough light for the short exposure time of each frame. Tracking markers were placed inside the scene and on the dynamic robot structure to facilitate more accurate COLMAP camera pose estimation.

	Synthetic Dataset 1 (Speaker)	Synthetic Dataset 2 (Robot)	Real-World Dataset (Robot)
Type of object dynamics	Sinusoidal membrane oscillation	Sinusoidal robot arm oscillation	Transient impact hammer excitation of the robot arm
Oscillation amplitude	+/- 1 mm	+/- 0.03 deg	no existing ground-truth

Oscilation amplitude ground-truth	-	+/- 3 deg	-
Oscillation period *T*	14 frames	25 frames	-
number of rendered / recorded frames (per view)	14 frames	50 frames	397 frames
framerate	25 fps	25 fps	640 fps
number of camera views (for dynamic scene)	24 views	3-50 views	27
number of additional camera views (for static scene)	-	100 views	-
Camera Pose Accuracy	ground-truth	ground-truth / perturbed poses	COLMAP pose estimates
Render Engine	Blender Eevee	Blender Eevee	-
Stage 1 MM-Factor	-	100x	75x

Table 1 Dataset Parameters

Figure 5 Synthetic Dataset 1 (Speaker) scene setup including evenly spaced cameras for maximal visual coverage via the Fibonacci Sphere algorithm [14] and exemplary training views.

Figure 6 Synthetic Dataset 2 (Robot) scene setup including evenly spaced cameras positioned via the Fibonacci Sphere algorithm and exemplary training views.

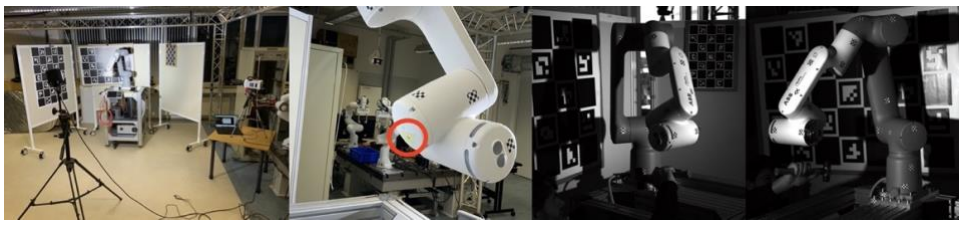

Figure 7 Real-World Dataset (Robot) scene setup with highlighted impact location in red, camera, laptop, LED light and exemplary training views.

4 Evaluation

To evaluate the method and different experimental parameters in our comprehensive synthetic scene setup we used 3 different loss metrics. L2 (Mean Squared Error) measures the average squared difference between pixels (lower is better). LPIPS (Learned Perceptual Image Patch Similarity) compares images based on perceptual similarity, considering how humans perceive differences (lower is better). PSNR (Peak Signal-to-Noise Ratio) measures the ratio of the maximum possible pixel value to the noise level (higher is better). 15 render-sequences from 15 test-views (not seen during training at Stage 2) were compared with their corresponding ground-truth sequences. For the ground-truth test-view sequences only the foreground is rendered, whereas for the rendered sequences from the trained model, a 3D bounding box is used to crop out all Gaussians outside the box to better evaluate the 3D reconstruction of the dynamic structure of interest. To obtain a final evaluation metric per experiment, all frames of a render-sequence (number of frames = number of test-views x number of timesteps) were compared against the ground-truth and averaged. It should be noted, that while the input training data of the method comprises rendered image-sequences with 1x motion amplitude (subsequently undergoing 100x Motion Magnification and Dynamic 3D Gaussian Splatting), the corresponding ground-truth test data comprises rendered image-sequences with native 100x motion amplitude with the rest of the parameters remaining the same.

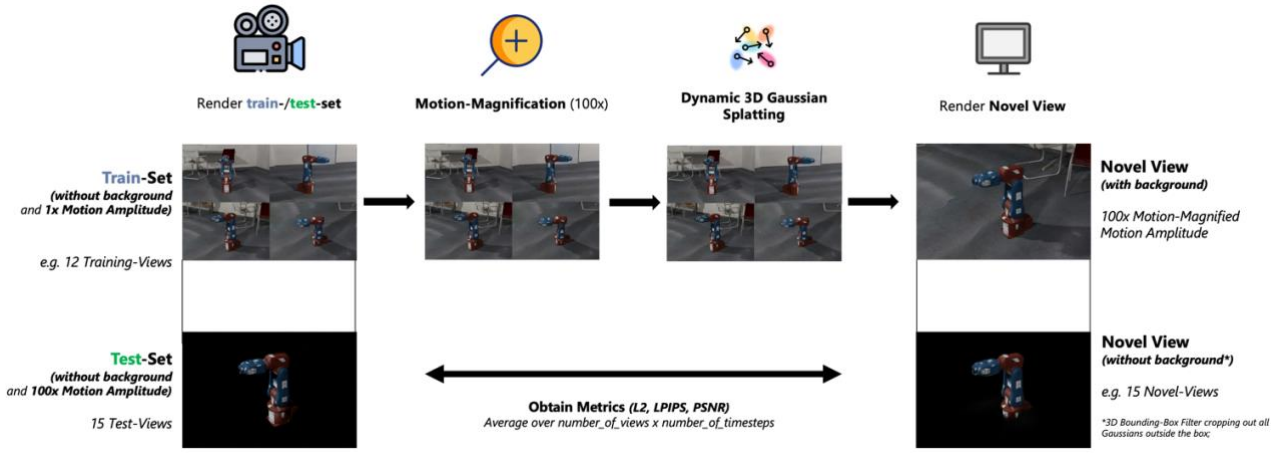

Figure 8 Evaluation of proposed method

5 Results and Discussion

This work introduces a novel method for visualizing and analyzing the 3D operational deflection shapes of mechanical systems using contactless, full-field measurements obtained through camera recordings. To the best of our knowledge, this is the first approach to combine learning-based Motion Magnification with Dynamic 3D Gaussian Splatting for this purpose. Our experiments, conducted on both synthetic and real-world data, demonstrate the effectiveness of the proposed method and its robustness to variations in experimental parameters.

5.1 Synthetic Experiments

5.1.1 Order of Stages

In our preliminary experiments with the simple Synthetic Dataset 1 (Speaker), the output quality of novel view renderings was qualitatively compared for the two different orderings of both stages: stage 1 (Motion Magnification) and stage 2 (Dynamic 3D Gaussians).

Figure 9 Left: First 3D Dynamic Gaussian Splatting, then Motion Magnification on Novel View Rendering.
Right: First Motion Magnification, then 3D Dynamic Gaussian Splatting.

Firstly, the order of performing the two stages Motion Magnification and Dynamic 3D Gaussian Splatting has a significant effect on the output quality. Optimizing a Dynamic 3D Gaussian Model on the raw input data without prior Motion Magnification is able to retain some of the subtle motions in 3D, but introduces artefacts and noise due to the dynamically moving 3D Gaussians, which is strongly amplified by the downstream Motion Magnification yielding poor results. Additionally, this ordering doesn't allow interactive rendering of novel views, since only the trained Dynamic 3D Gaussian Model is real-time performant. Therefore, applying the Motion Magnification in the first stage prior to optimizing the Dynamic 3D Gaussian Model in the second stage is the preferred order and was maintained for all consecutive experiments.

5.1.2 Number of training views

Our sensitivity analysis regarding the number of training views indicates that 8-15 equidistantly placed training views provide adequate coverage for a half-hemisphere centered around the object as can be seen in **Error! Reference source not found.** Figure 10 and Figure 11**Error! Reference source not found.**. Although additional views can further enhance the 3D reconstruction quality, the improvement is marginal. By initializing the DGS model (stage 2) with a comprehensive scan (e.g., 100 images) at the first timestep before excitation, we can significantly enhance the quality with minimal additional experimental cost as seen in Figure 21. Acquiring 100 images of a static scene is experimentally much simpler than capturing 100 image sequences of a dynamic scene, which would require either 100 cameras recording simultaneously or one camera recording the dynamic scene asynchronously 100 times. While the first approach is expensive to set up, the latter approach introduces potential experimental errors due to non-identical dynamic behavior between recordings and increases the time cost of the experiment.

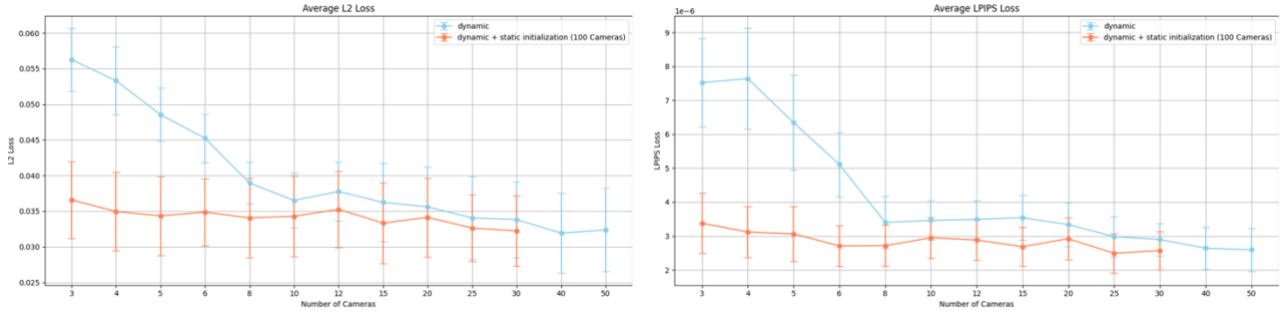

Figure 10 Average L2/LPIPS Loss and standard deviation for 13 experiments with varying number of training views + 11 experiments with an additional 100 training-views for initialization at first timestep

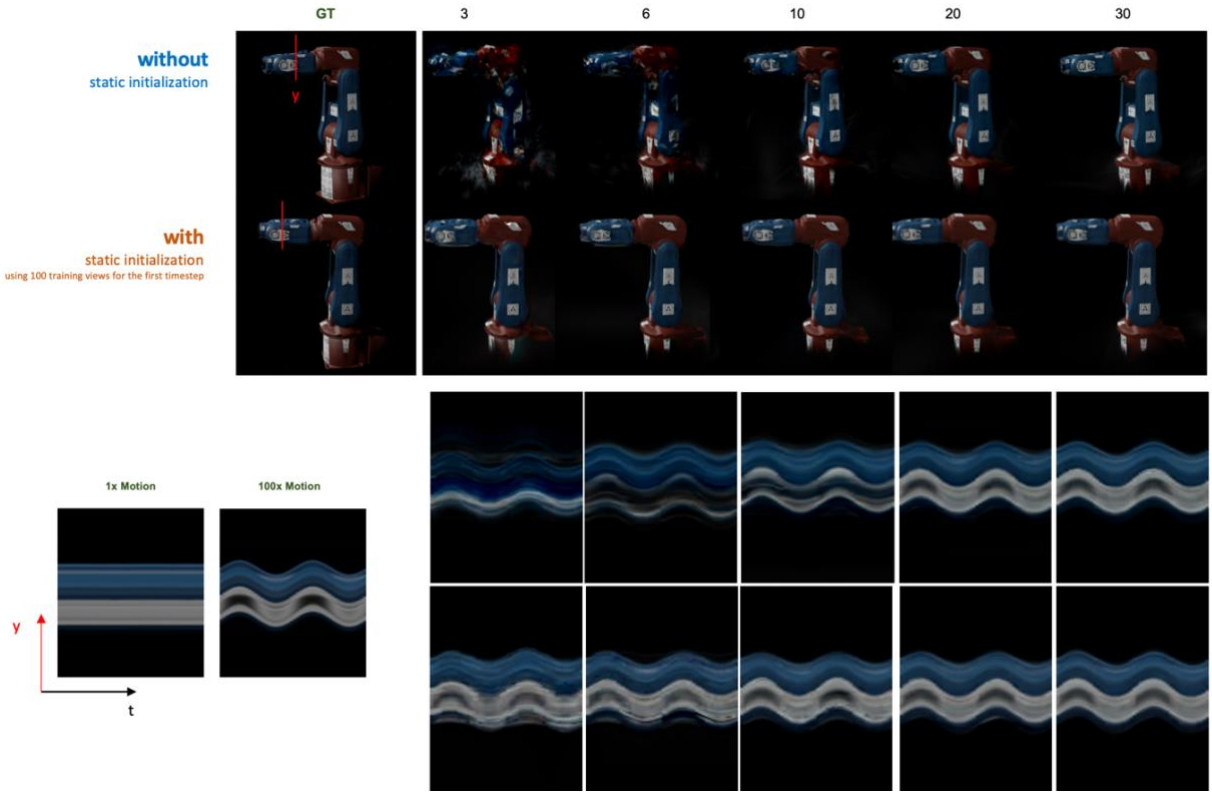

Figure 11 Average L2/LPIPS Loss and standard deviation for varying number of training views (+ conditioning on 100 additional training-views at first timestep).
The red line y denotes the slice that is plotted over time.

5.1.3 Camera-Pose Accuracy

In our sensitivity analysis regarding the provided camera pose accuracy for the DGS model (stage 2) the camera extrinsic was perturbed by some Gaussian Noise to measure the robustness against camera pose estimation errors. While the pipeline is somewhat resilient to low pose noise levels, higher noise levels result in noticeable image and motion blurring as can be seen in the Y-T slices in **Figure 13**. One insightful observation from the "Number of Training-Views" experiments is the significant improvement in reconstruction quality achieved by initializing the Dynamic 3D Gaussian Model with a comprehensive capture of the static scene at the first timestep. This initialization not only helps provide a better 3D reconstruction with less training views of the dynamic scene, but these additional views also help refine the camera pose accuracy of the training views of the dynamic scene during the pose estimation step with COLMAP which directly benefit the subsequent dynamic reconstruction. This underscores the importance of strategies that improve camera pose accuracy, such as using a comprehensive static scene initialization, to achieve optimal dynamic reconstruction quality, especially in real-world scenarios where pose estimation errors are inevitable.

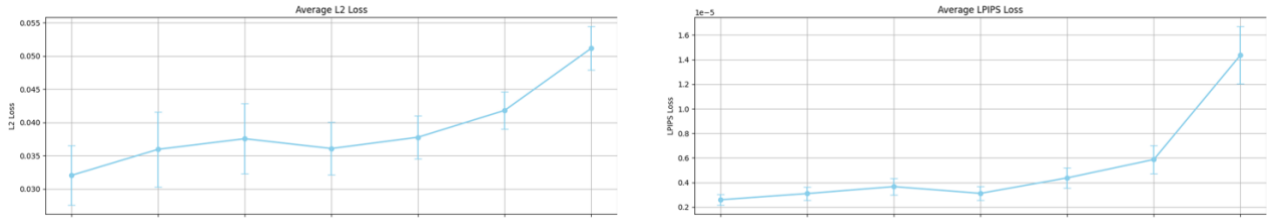

Figure 12 Average L2/LPIPS Loss and standard deviation for different camera pose noise levels

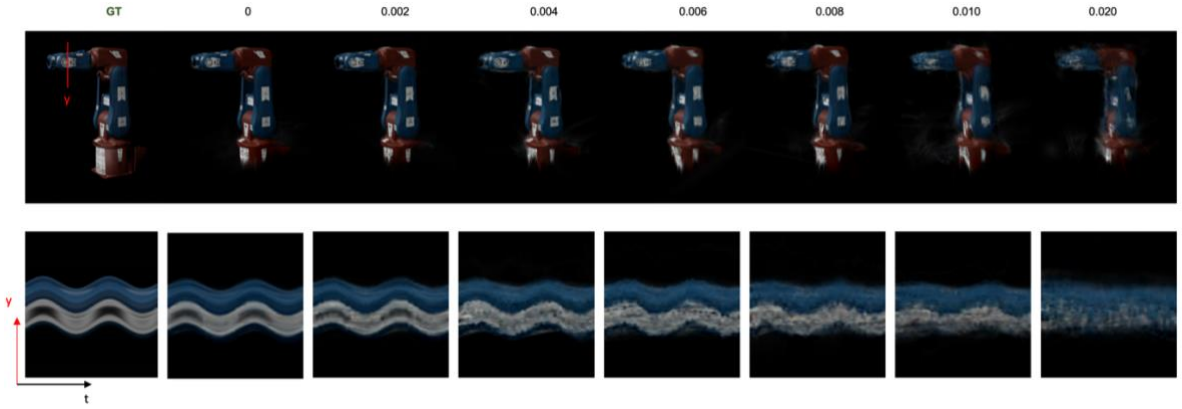

Figure 13 Comparison images and temporal y-t slices for different pose noise levels of up to 2cm std.

5.1.4 Camera RGB-Noise

Real-world images often suffer from noise due to sensor limitations and environmental conditions. Our experiments with synthetically added RGB noise reveal the limitations of the pipeline's robustness to high noise levels. While the method performs well with low levels of noise, the results degrade noticeably as the noise level increases, even with temporal filtering applied during the Motion Magnification stage. It mitigates the effects of noise to some extent, but it requires careful tuning of the bandpass filter frequencies (low and high cutoff frequencies), which demands a certain level of expertise and might not be universally applicable. The real-world experiment, which utilized a high-speed camera operating at a fast frame rate and a small aperture, highlights the importance of minimizing noise at the source. By providing sufficient lighting and using appropriate camera settings, we can ensure that the input image sequences have low levels of RGB noise, facilitating more accurate motion magnification and better overall reconstruction quality.

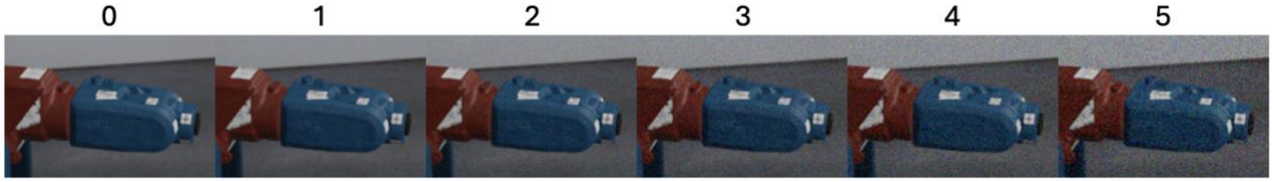

Figure 14 Comparison of different RGB-Noise Levels in input training view.

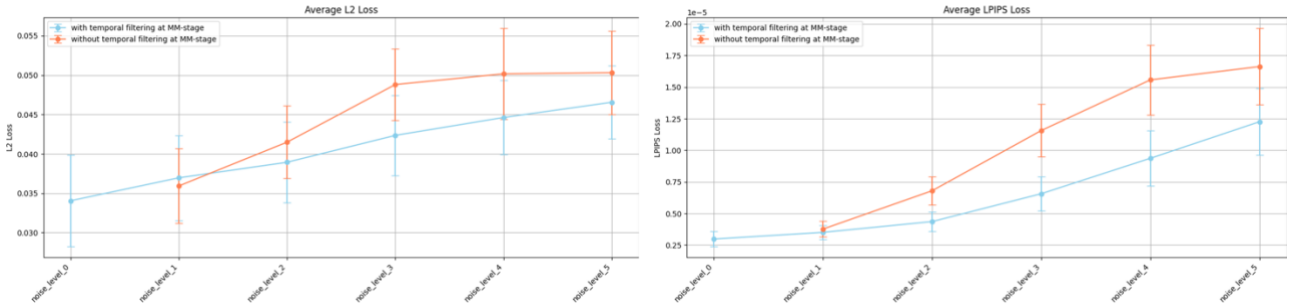

Figure 15 L2 and LPIPS metrics for RGB Experiments

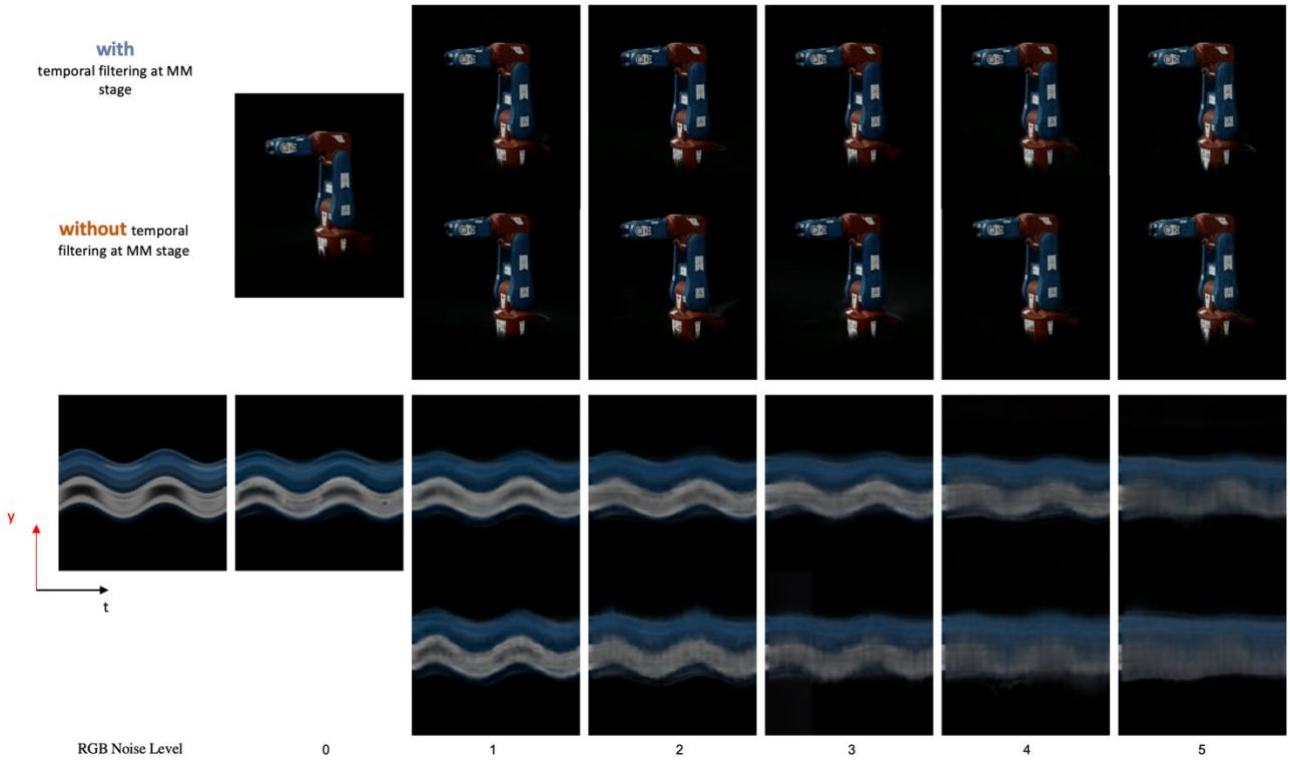

Figure 16 Comparison images and temporal Y-T Slices for noise levels 0.01 to 0.05.

5.2 Real-World Experiment

To validate the efficacy of the method in the real world, the GoFA CRB15000 Robot was used as the dynamic structure and excited by an impact hammer. The Dynamic 3D Gaussian Model was trained on 23 training views without static initialization and novel-view renderings from our method were compared to a simple Motion- Magnification of the Ground-Truth Image-Sequence to assess, whether the reconstruction quality of our method degenerates the visual result significantly compared to just peforming motion magnification. A motion magnification factor of 75x was used in this experiment. Due to time-constraints, no accelerometer measurements were performed for the real-world example. Thus, leaving an interesting path for further research in this area. Link to videos: https://syncandshare.lrz.de/getlink/fi8Gd1oxcjjc5q2X51xfvA/.

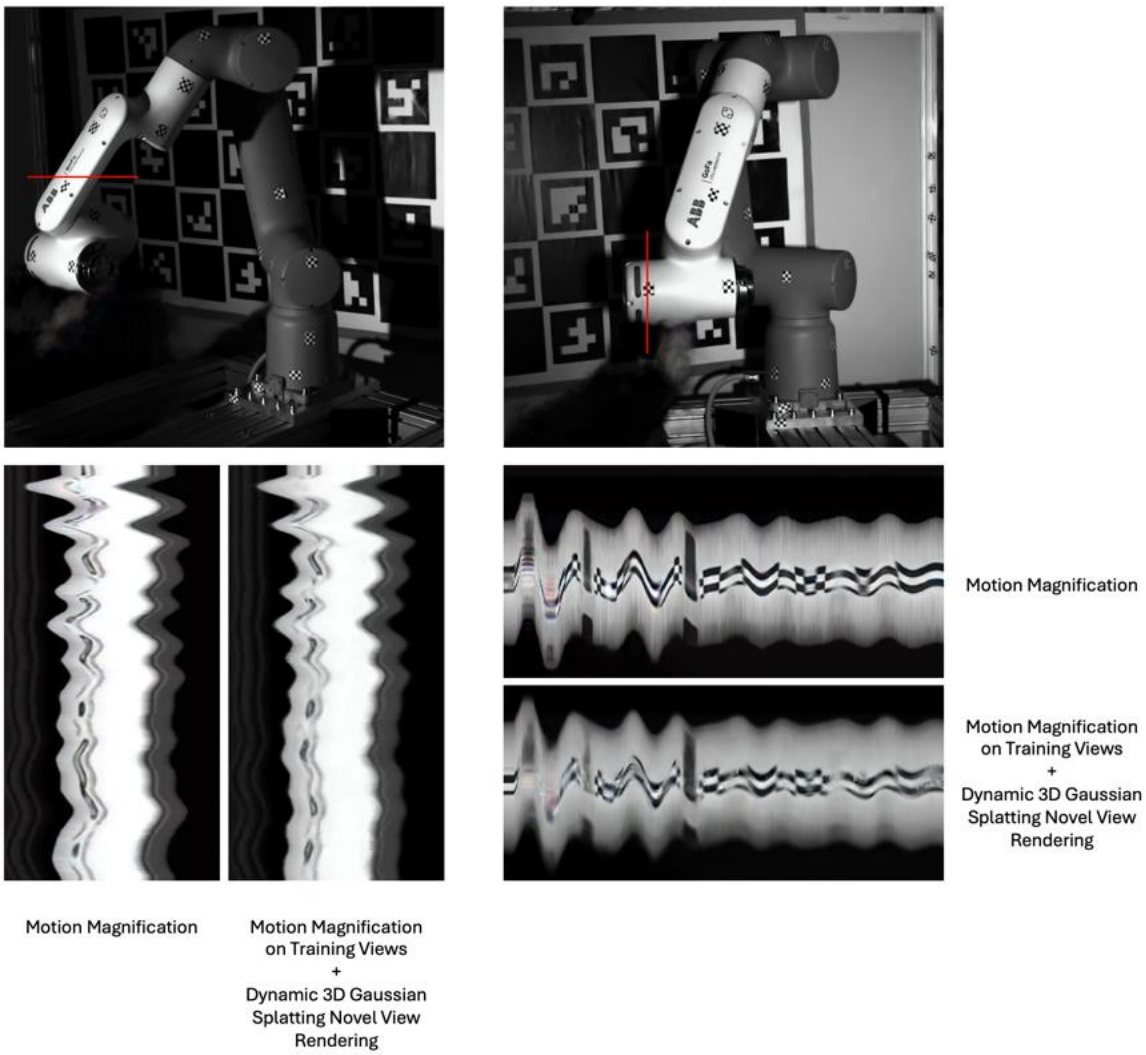

Figure 17 Top: motion magnified test views. Bottom: x-t slices and y-t slices for both test views.

6 Conclusion

This work introduces a novel method for visualizing and analyzing the 3D operational deflection shapes (ODS) of mechanical systems using a combination of learning-based Motion Magnification and Dynamic 3D Gaussian Splatting. By amplifying subtle, often imperceptible, vibrations and reconstructing them in 3D, the pipeline offers a contactless, full-field approach to analyzing structural dynamics, providing a more intuitive and interactive means of understanding complex mechanical behavior.

To the best of our knowledge, this is the first approach to combine Motion Magnification with Dynamic 3D Gaussian Splatting for visualizing ODS. Our experiments, conducted on both synthetic and real-world data, demonstrate the effectiveness of the proposed method and its robustness to variations in experimental parameters, such as camera viewpoint, noise levels, and pose estimation accuracy.

While this method does not directly determine modal parameters like natural frequencies, mode shapes, or damping ratios, as provided by traditional Experimental Modal Analysis (EMA), it offers a valuable complementary tool for engineers. The ability to visualize ODS in 3D, without the need for physical sensors or artificial excitation, can provide unique insights into the dynamic behavior of structures under operational conditions. The proposed method could therefore be used as a preliminary step to determine the correct placement of acceleration sensors for a subsequent EMA to capture the most dynamical parts of a structure. Alternatively, the visual representation of the operational motion could be used to check the validity of EMA extracted mode shapes. Finally, this method could be applied for quick non-intrusive inspection of complex systems to detect faults without requiring expert understanding of a system or modal analysis tools.

7 Future Work

Future research could explore several avenues to further enhance this method. One direction would be to investigate the integration of accelerometer or force measurements. This could involve using the data to validate the visualized ODS or even combining the data in a hybrid approach to improve the accuracy and detail of the 3D reconstructions.

Additionally, exploring ways to extract quantitative data from the 3D reconstructions could be valuable. This might involve developing methods to accurately determine the Cartesian movement of surface points from the 3D Gaussian representation, potentially leading to the estimation of some modal parameters. However, even with such advancements, this approach would likely complement rather than replace traditional EMA techniques.

Finally, optimizing the algorithms for Motion Magnification and Dynamic 3D Gaussian Splatting would reduce the required processing time, making the method even more practical for real-time applications and analysis of more complex mechanical systems. Improving the robustness of the method to poorer quality inputs could facilitate the use of alternative acquisition systems such as smartphones. This would enable faster, cheaper and easier application of the method.

References

[1] T.-H. Oh *et al.*, “Learning-based Video Motion Magnification.” arXiv, Jul. 31, 2018. Accessed: May 30, 2024. [Online]. Available: http://arxiv.org/abs/1804.02684

[2] J. Luiten, G. Kopanas, B. Leibe, and D. Ramanan, “Dynamic 3D Gaussians: Tracking by Persistent Dynamic View Synthesis.” arXiv, Aug. 18, 2023. Accessed: May 30, 2024. [Online]. Available: http://arxiv.org/abs/2308.09713

[3] R. Allemang and P. Avitabile, *Handbook of Experimental Structural Dynamics*. Springer New York, 2013. [Online]. Available: https://books.google.co.uk/books?id=exKizgEACAAJ

[4] Siemens, “Modal Testing: A Practical Guide.” [Online]. Available: https://community.sw.siemens.com/s/article/Modal-Testing-A-Guide

[5] D. J. Ewins, *Modal testing : theory, practice, and application*, 2nd ed., 1 online resource (xiii, 562 pages) : illustrations vols. in Mechanical engineering research studies; 10. Baldock, Hertfordshire, England: Research Studies Press, 2000. [Online]. Available: http://catalog.hathitrust.org/api/volumes/oclc/37928483.html

[6] Siemens, “What is an Operational Deflection Shape (ODS)?” [Online]. Available: https://community.sw.siemens.com/s/article/what-is-an-operational-deflection-shape-ods

[7] K. O’Shea and R. Nash, “An Introduction to Convolutional Neural Networks.” arXiv, Dec. 02, 2015. Accessed: Jun. 17, 2024. [Online]. Available: http://arxiv.org/abs/1511.08458

[8] A. F. Agarap, “Deep Learning using Rectified Linear Units (ReLU).” arXiv, Feb. 07, 2019. Accessed: Jun. 17, 2024. [Online]. Available: http://arxiv.org/abs/1803.08375

[9] B. Kerbl, G. Kopanas, T. Leimkühler, and G. Drettakis, “3D Gaussian Splatting for Real-Time Radiance Field Rendering.” arXiv, Aug. 08, 2023. Accessed: May 30, 2024. [Online]. Available: http://arxiv.org/abs/2308.04079

[10] K. Čufar, J. Slavič, and M. Boltežar, “Mode-shape magnification in high-speed camera measurements,” *Mechanical Systems and Signal Processing*, vol. 213, p. 111336, 2024, doi: https://doi.org/10.1016/j.ymssp.2024.111336.

[11] Blender Foundation, “Blender.” 2024. [Online]. Available: https://www.blender.org

[12] Johannes Lutz Schönberger and Jan-Michael Frahm., “Structure-from-Motion Revisited,” *Conference on Computer Vision and Pattern Recognition (CVPR). 2016.*.

[13] Johannes Lutz Schönberger and Jan-Michael Frahm, “Pixelwise View Selection for Unstructured Multi-View Stereo,” *European Conference on Computer Vision (ECCV). 2016.*.

[14] Á. González, “Measurement of Areas on a Sphere Using Fibonacci and Latitude–Longitude Lattices,” *Mathematical Geosciences*, vol. 42, no. 1, pp. 49–64, Jan. 2010, doi: 10.1007/s11004-009-9257-x.