Vibration Frequency Estimation via AI-Based Motion Magnification

25-1 DLIP Final Project Proposal

School of Mechanical and Control Engineering

Prof. YoungKeun-Kim

22000355 HyeonGyu-Seo 22200757 Joungbin Choi

한동대학교

Introduction

Machine Defects

1. Condition monitoring of existing industrial facilities is performed based on physical sensors.

2. This project aims to implement non-contact sensing technology

3. Using Motion Magnification to amplify micro-vibrations in images

Problem Statement

1. Situations where physical sensor attachment is difficult
   - high temperature/high pressure environment, narrow space, moving objects, etc.

2. Sensor environment vulnerable to noise - magnetic field interference, structural vibration confusion

3. When single component vibration localization is required
   - status analysis of specific parts, not the entire facility

Goal of Project & Target Customers

The Main Goal

To solve the problems mentioned in the problem statement, we want to analyze vibrations using cameras and deep learning models. This will allow us to predict the possibility of failure and prevent plant losses.

Target Performance

- **Model**
  - PSNR > 25.0 dB
  - SSIM > 0.9
- **Visualization & Detect Vibration**
  - XT-Slice Visualization
  - Image Processing

Target Customers

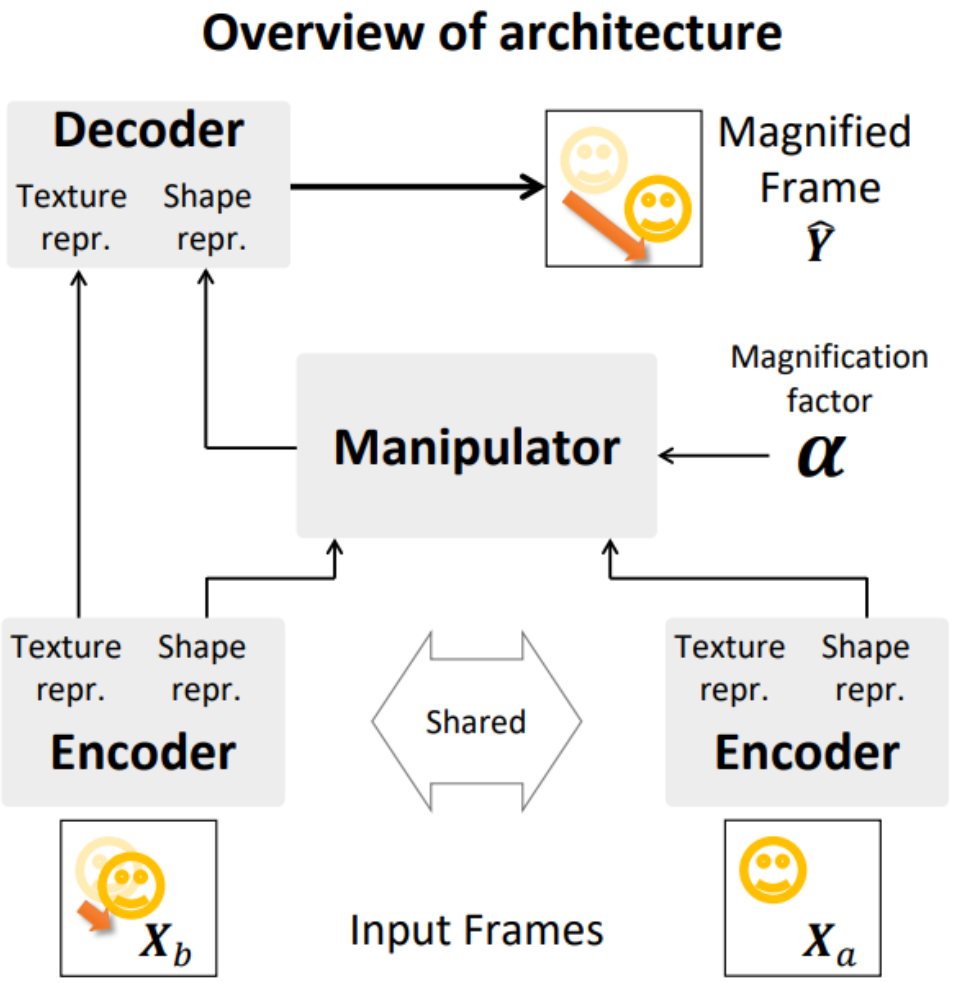
1. Smart Factory Integrators – for condition monitoring of machines in automated manufacturing environments
2. Plant Maintenance Service Providers – where large-scale equipment needs frequent non-invasive inspection
3. Robot Health Diagnostics Companies – enabling remote or visual-based state estimation for robots or mechatronic systems

Magnification Model Architecture

1. Encoder
   - Decompose input frame into two components
   - Shape Representation (M)
   - Texture Representation (V)

2. Manipulator
   - non-linear processing via convolution
   - $G_m(M_a, M_b, \alpha) = M_a + h\big(\alpha \cdot g(M_b - M_a)\big)$

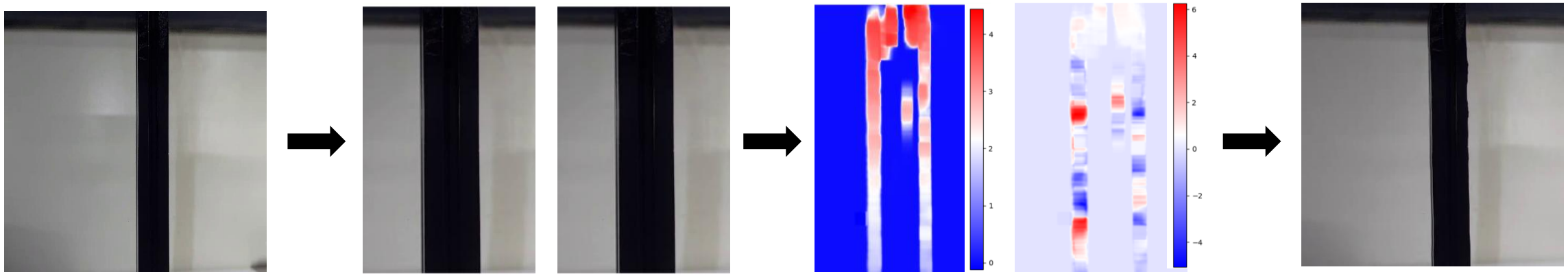
3. Decoder
   - Combines the manipulated shape and texture.
   - Reconstructs the magnified output frame.



Create Train Dataset

1. Extract frames from video
2. Construct input pairs (Xa, Xb)
3. Estimate pixel movement between two consecutive frames
4. Generate target (Y) using Optical Flow warping

여기다 실험 세팅 한 거 사진 하나?



Input video.mp4 | Xa(F_{prev}) | Xa(F_{curr}) | dx(Horizontal) | dy(Vertical) | Warped Y (x 10.0)

Evaluation

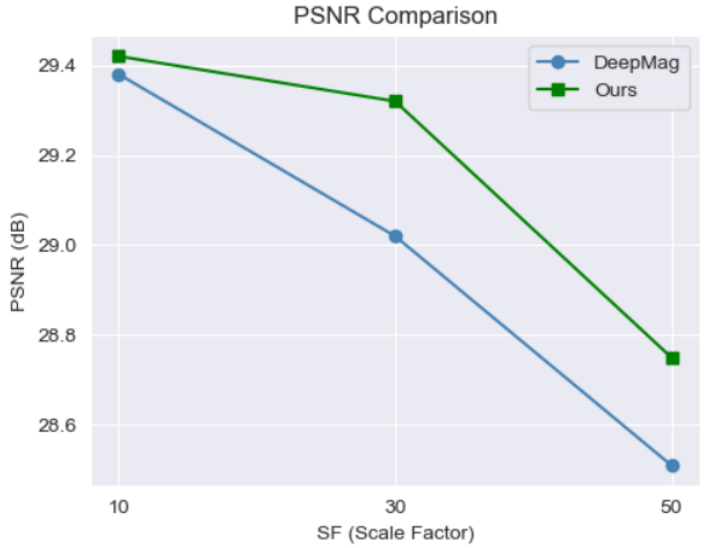
1. PSNR (Peak Signal-to-Noise Ratio)
   - $MSE = \frac{\Sigma_{M,N}[I_1(m,n)-I_2(m,n)]^2}{M*N}$
   - $PSNR = 10\log_{10}\left(\frac{R^2}{MSE}\right)$
   - The loss information on the image quality
   - The higher the value, the less the loss.

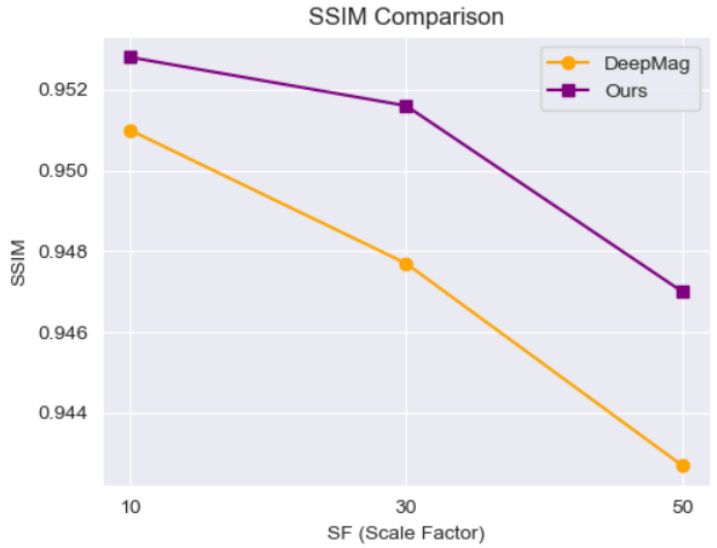
2. SSIM (Structural Similarity Index Map)
   - Evaluates quality in three aspects: Luminance, Contrast, and Structural.
   - $SSIM(x,y) = [I(x,y)]^{\alpha} \cdot [c(x,y)]^{\beta} \cdot [s(x,y)]^{\gamma} = \frac{(2\mu_x\mu_y+C_1)(2\sigma_{xy}+C_2)}{(\mu_x^2+\mu_y^2+C_1)(\sigma_x^2+\sigma_y^2+C_2)}$

Evaluation

Ground Truth	Deepmag (SF=10)	Deepmag (SF=30)	Deepmag (SF=50)
This image	29.38/0.9510	29.02/0.9477	28.51/0.9427
Set 10 means	24.66/0.9318	24.61/0.9256	23.98/0.9129

Ground Truth	Ours (SF=10)	Ours (SF=30)	Ours (SF=50)
This image	29.42/0.9528	29.32/0.9516	28.75/0.9470
Set 10 means	25.05/0.9362	24.66/0.9329	24.22/0.9252





Evaluation

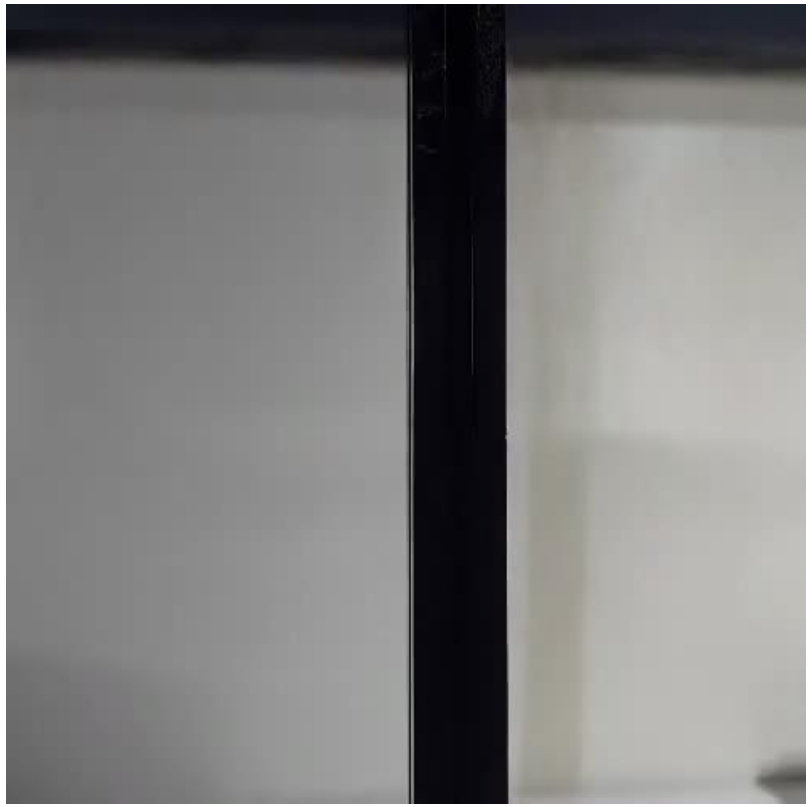


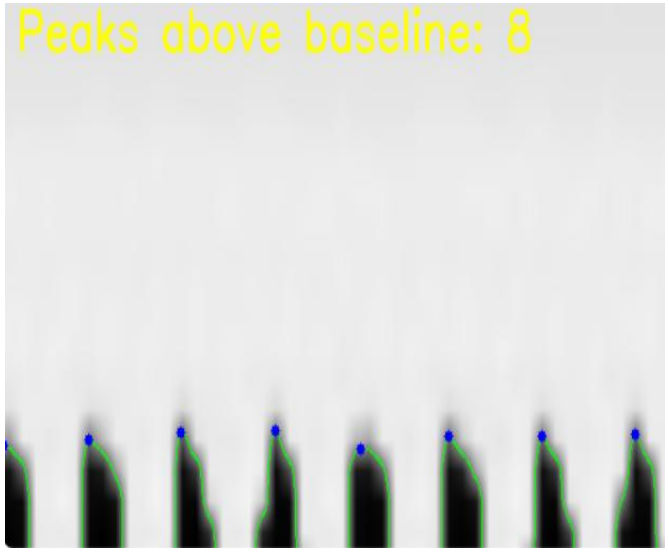
Image processing

1. Create X-T (Horizontal)

2. Periodicity of Slice Image → Image-based Vibration Frequency Estimation Possible
   - Display analysis coordinates as lines for each frame
   - Find contours in the edge image and consider each as a "vibration waveform".

Magnified Video

X-T slice Video



Magnified Video

Result