Multimodal sensor-guided diffusion model for machined surface image synthesis

Jae Gyeong Choi
Ulsan National Institute of Science and Technology
Ulsan, Republic of Korea
choil6043@unist.ac.kr

Sunghoon Lim
Ulsan National Institute of Science and Technology
Ulsan, Republic of Korea
sunghoonlim@unist.ac.kr

ABSTRACT

Generative models, particularly diffusion-based approaches, have gained significant attention in recent years due to their ability to create realistic outputs. Despite their potential, the application of these models in manufacturing remains largely unexplored. This work presents a framework that addresses this gap by generating machined surface images guided by multiple sensor inputs in manufacturing. The proposed model integrates information from multiple sensors with varying sampling rates using multimodal embedding and employs a latent diffusion model to translate the fused sensor embedding into an image embedding, which is then converted into a machined surface image. The effectiveness of the framework is validated using real-world time-series data, including force, torque, acceleration, sound, voltage, and current, collected from a carbon-fiber-reinforced plastic drilling process. The results demonstrate the model's ability to predict delamination from the generated machined surface images. The proposed approach has the potential to enhance process monitoring, quality control, and predictive maintenance in smart manufacturing by enabling sensor-guided visual inspection and defect detection.

CCS CONCEPTS

• **Applied computing** → Industry and manufacturing; • **Computing methodologies** → Machined surface image generation.

KEYWORDS

Image synthesis; sensor-to-image; generative model; latent diffusion model

ACM Reference Format:

Jae Gyeong Choi and Sunghoon Lim. 2018. Multimodal sensor-guided diffusion model for machined surface image synthesis. In *Proceedings of (Conference acronym 'XX).* ACM, New York, NY, USA, 2 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

Recent advancements in generative models, particularly diffusion-based approaches, have led to significant breakthroughs in various domains, including text-to-image generation [3, 4]. While these models have demonstrated remarkable performance in creating realistic images from textual descriptions, their potential in manufacturing applications has not been fully explored. In manufacturing sites, the ability to generate accurate visual representations of machined surfaces based on sensor data can greatly benefit process monitoring, and predictive maintenance.

In this work, we propose Sensor2Image++, a framework for generating machined surface images guided by multimodal sensor inputs. Building upon the success of our previous work, Sensor2Image [2], which translates single sensor data into images, Sensor2Image++ excels in synthesizing high-fidelity machined surface images while effectively addressing the challenge of integrating information from multiple sensors with varying sampling rates. This enhancement ensures that our model comprehensively captures the intricacies of the manufacturing process, thereby advancing the state of the art in sensor-guided image synthesis.

2 METHOD

2.1 Latent diffusion model for machined surface image synthesis

The proposed approach employs a latent diffusion model architecture for generating images of machined surfaces. The model is constructed upon a variational autoencoder (VAE) framework, which comprises an encoder and a decoder. The input image of the machined hole surface, x_0, is fed into the encoder $E_\theta(\cdot)$, which maps the input to a latent representation z_0 in the latent space $\mathcal{Z}$. To generate a single machined surface image corresponding to multiple sensor inputs collected from a process, it is essential that the generative model is capable of producing deterministic outputs. To achieve this, we employ the dnoising diffusion implicit models (DDIM) sampling process [5]. DDIM introduces a deterministic schedule for the denoising process, allowing for the generation of consistent and reproducible images. The DDIM sampling process is defined by a sequence of latent variables , denoted by $z_1, z_2, \ldots, z_T$, where T is the total number of diffusion steps. The latent variable z_t is obtained by iteratively applying the DDIM update rule:

$$z_{t-1} = \sqrt{\bar{\alpha}_{t-1}} \left(\frac{z_t - \sqrt{1-\bar{\alpha}_t}\epsilon_\theta(z_t, t)}{\sqrt{\bar{\alpha}_t}} \right) + \sqrt{1-\bar{\alpha}_{t-1}}\epsilon_\theta(z_t, t) \quad (1)$$

where $\bar{\alpha}_t$ is a deterministic variance schedule, and $\epsilon_\theta(z_t, t)$ is a learned denoising function that predicts the noise added at each step.

The decoder $D_\phi(\cdot)$ takes the noisy latent representation z_T and reconstructs the generated image $\hat{x}_0 = D\phi(z_T)$. By training the model to denoise the latent space, the underlying structure and characteristics of the machined surface images are effectively captured, enabling the generation of realistic and diverse samples.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

Conference acronym 'XX, June 03–05, 2018, Woodstock, NY

© 2018 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 978-1-4503-XXXX-X/18/06
https://doi.org/XXXXXXX.XXXXXXX