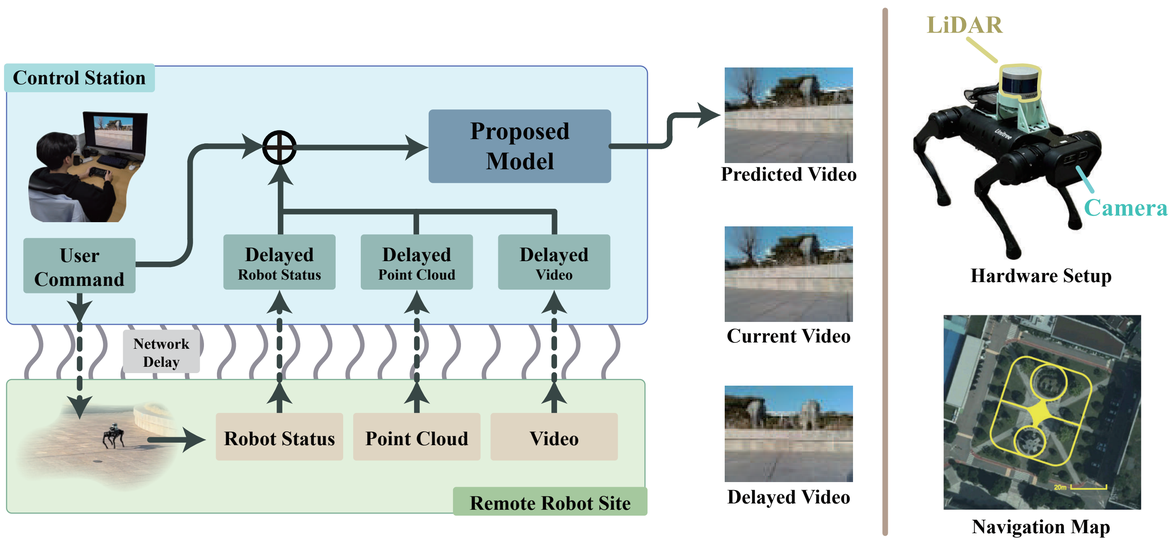
# Prediction of Delay-Free Scene for Quadruped Robot Teleoperation: Integrating Delayed Data with User Commands

Seunghyeon Ha, Seongyong Kim, and Soo-Chul Lim
Dongguk University, Seoul, Republic of Korea

Page + Videos

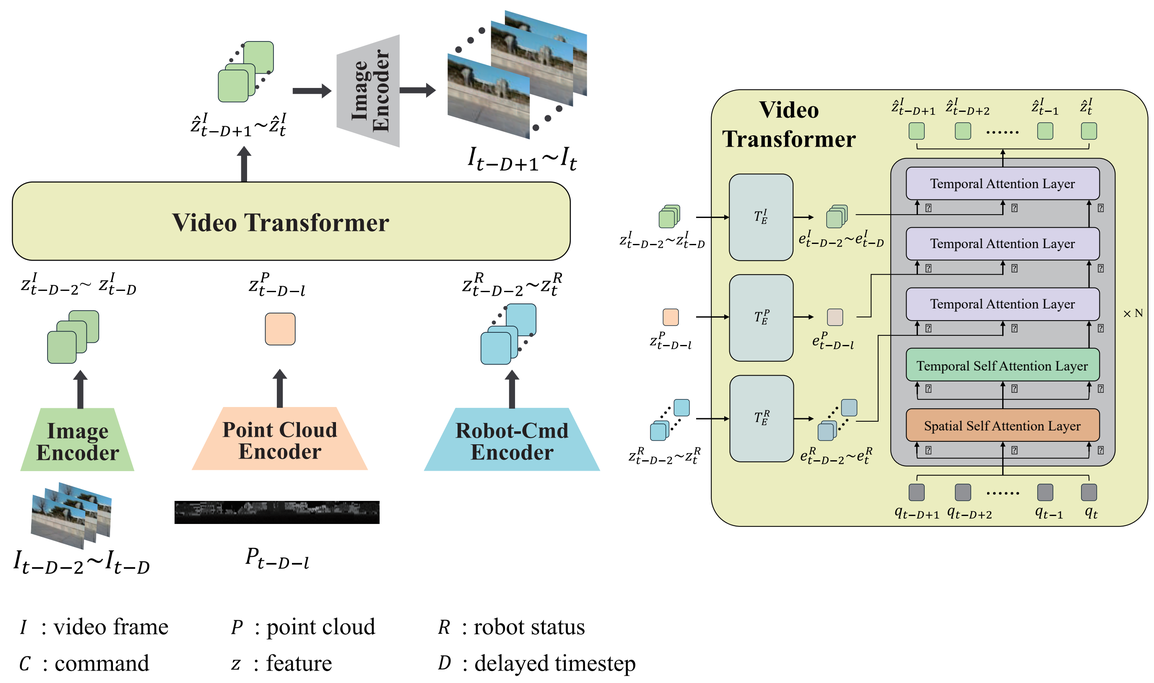
**This paper addresses the problem of delayed visual feedback in quadruped robot teleoperation caused by network latency, and proposes a transformer-based video prediction model that integrates delayed image, LiDAR point clouds, and robot status with real-time user commands to generate delay-free scene**

## Overview




- Focuses on visual delay in teleoperation under network latency
- Utilize delayed data and user command to generate delay-free scene

## Network Architecture

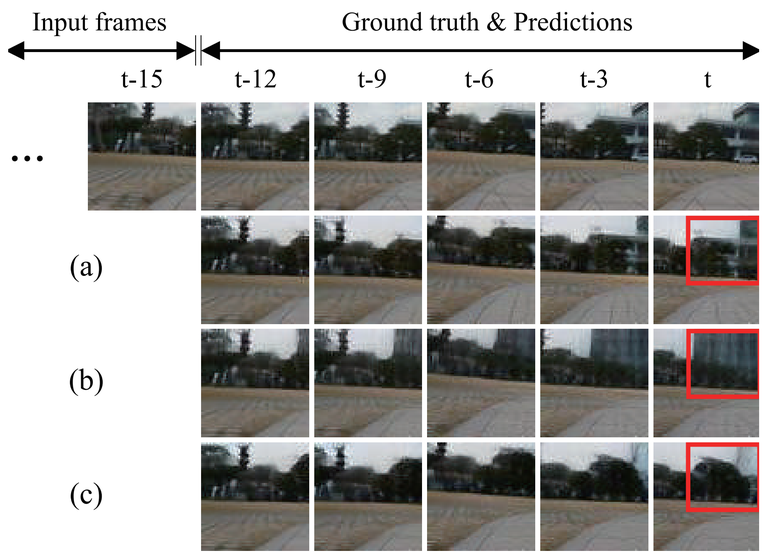



$I$ : video frame $P$ : point cloud $R$ : robot status
$C$ : command $z$ : feature $D$ : delayed timestep

- Pretrained *Encoders* extract data features
- *Video Transformer* predicts the feature of the current frame using the extracted features
- *Image Decoder* outputs the frame features to RGB images

## Ablation Study

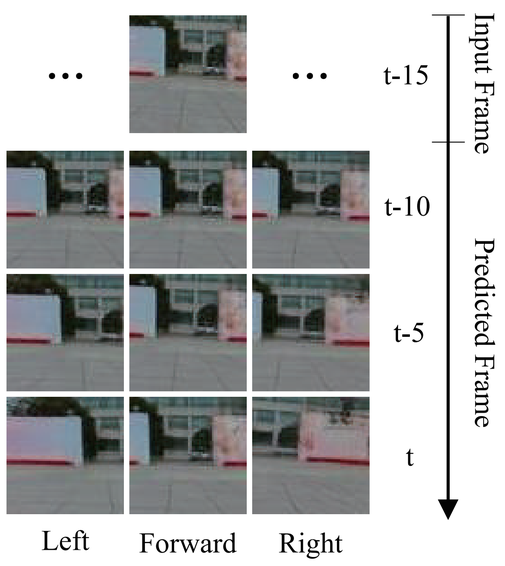
A. Effect of Point Cloud Data




(a) the *original* data
(b) the *maximum* values
(c) the *minimum* values

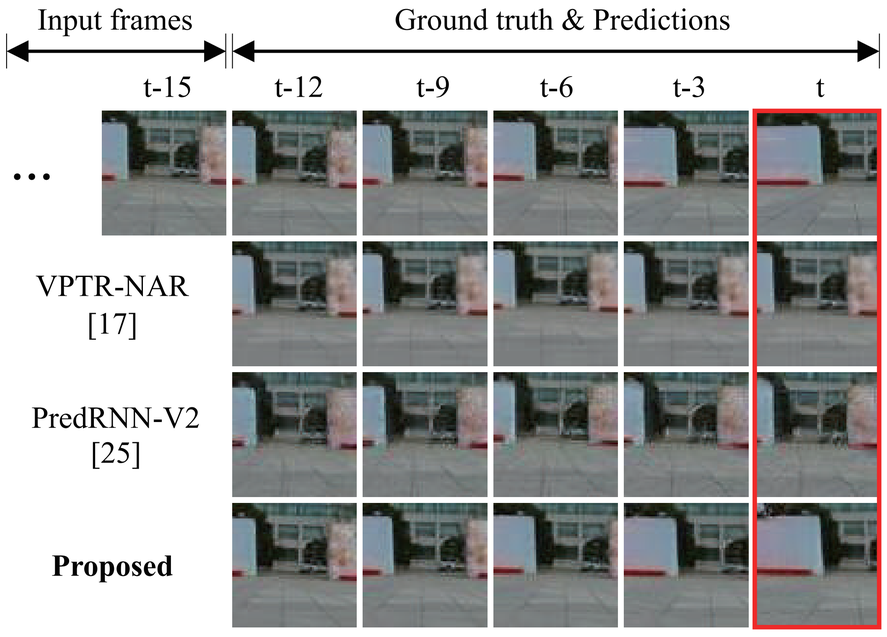
- Minimum-values denote unmeasurable regions, e.g., the sky.
- Maximum-values caused incorrect building predictions.
- Scene prediction results vary with changes in the point cloud input

B. Effect of Command Data




- Each prediction is the result of using the same input frame and different commands
- The proposed model reflects input commands by correctly predicting scenes

## Comparison of Qualitative Result




- Scenario: the robot turns left after moving straight forward
- Other models only predicted the robot moving straight
- The proposed model can generate turning scenes based on commands

TABLE: Comparison of qualitative result with other models

| | RMSE(↓) | PSNR(↑) | SSIM(↑) | LPIPS(↓) |
|---|---|---|---|---|
| Proposed | **0.0132** | **20.5956** | **0.6117** | **0.0684** |
| VPTR-NAR [17] | 0.0188 | 18.8938 | 0.5369 | 0.2336 |
| PredRNN-V2 [25] | 0.0207 | 18.4134 | 0.4954 | 0.1138 |