INTERNATIONAL INSTITUTE OF INFORMATION TECHNOLOGY
HYDERABAD

R&D SHOWCASE 2021

# DeepMPCVS: Deep Model Predictive Control for Visual Servoing

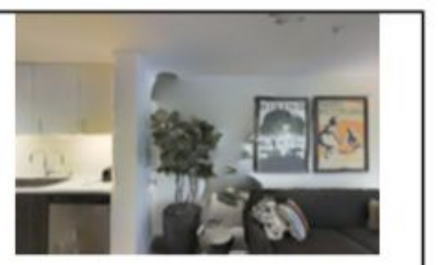
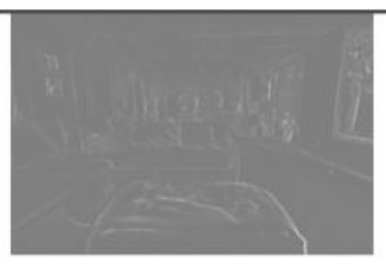
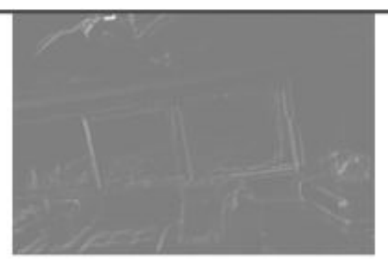
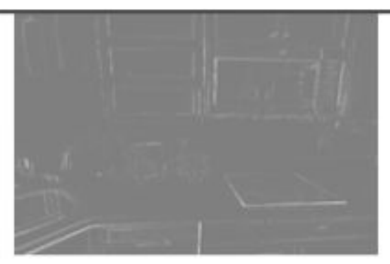
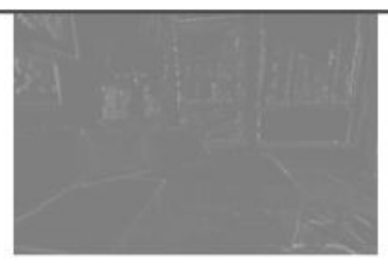
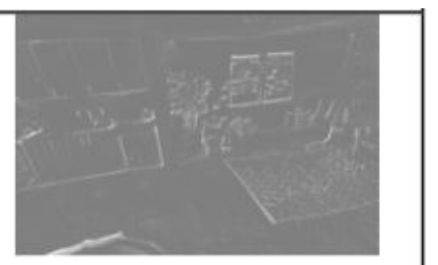
Robotics Research Center

## ABSTRACT

The simplicity of the visual servoing approach makes it an attractive option for tasks dealing with vision-based control of robots in many real-world applications. However, attaining precise alignment for unseen environments pose a challenge to existing visual servoing approaches. While classical approaches assume a perfect world, the recent data-driven approaches face issues when generalizing to novel environments. In this paper, we aim to combine the best of both worlds. We present a deep model predictive visual servoing framework that can achieve precise alignment with optimal trajectories and can generalize to novel environments. Our framework consists of a deep network for optical flow predictions, which are used along with a predictive model to forecast future optical flow. For generating an optimal set of velocities we present a control network that can be trained on-the-fly without any supervision. Through extensive simulations on photo-realistic indoor settings of the popular Habitat framework, we show significant performance gain due to the proposed formulation vis-a-vis recent state of the art methods. Specifically, we show a faster convergence and an improved performance in trajectory length over recent approaches.

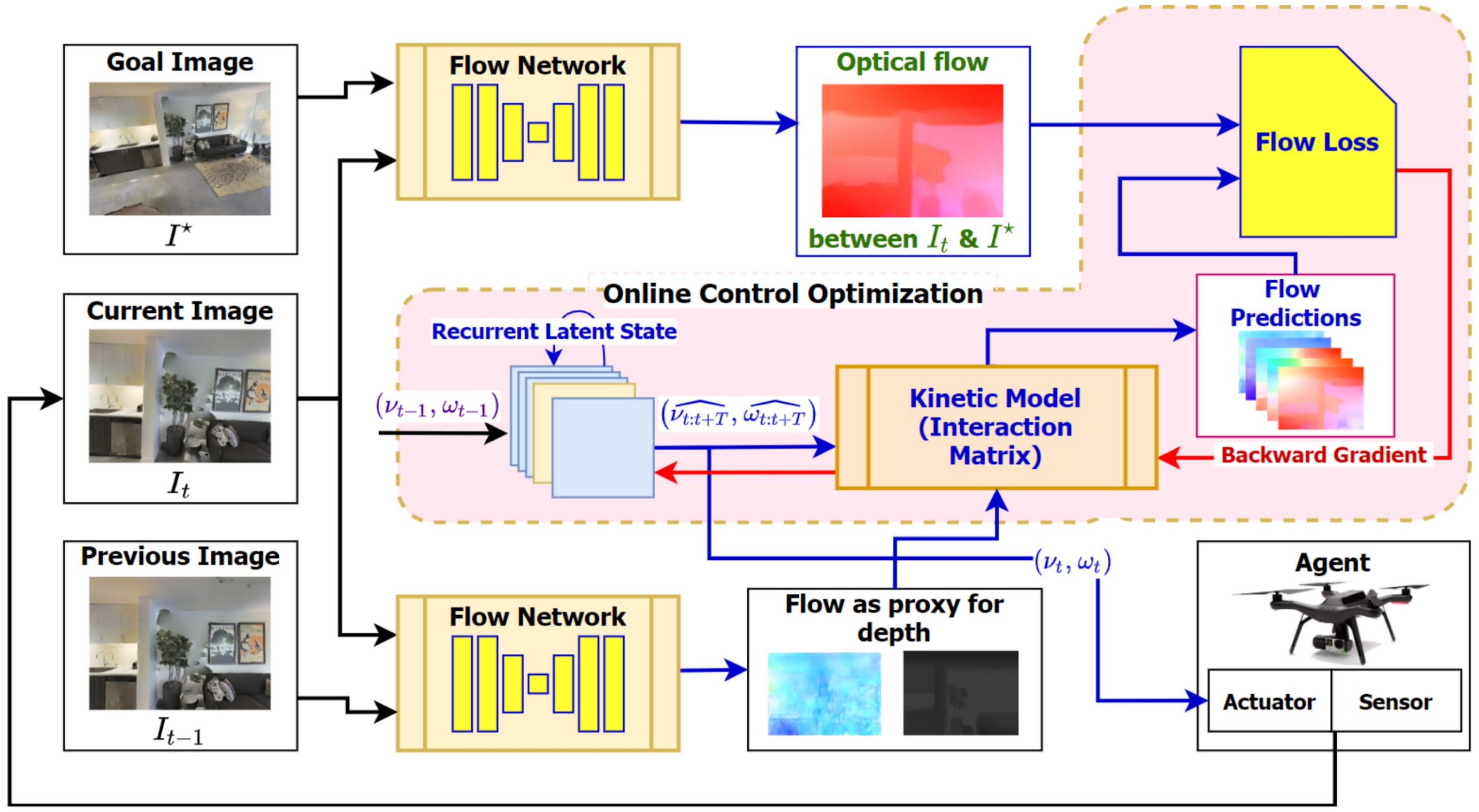
## SIMULATION RESULTS ON BENCHMARK

Our benchmark comprises of 10 indoor photo-realistic environments from the Gibson dataset in the Habitat simulation engine.. We use a free-flying RGB camera as our agent so that agent can navigate in all 6-DoFs without any constraint. We show the photometric error image representation computed between the desired image *I∗* and the final image on termination. Our convergence criteria for the run is photometric error of 500 or less.

| | | | | | |
|---|---|---|---|---|---|
| Initial Image | | | | |  |
| Desired Image | | | | | |
| MPC + LSTM [Ours] |  |  |  |  |  |

## MOTIVATION AND OBJECTIVE

Given agent observation *It* and desired observation *I∗* the pipeline generates optimal control commands [*vt* , *ωt*] for each time-step t until convergence is achieved. The flow network encodes optical flow between *It* and *I∗* and the control-optimizer architecture learns to generate optimal control for each time-step *t* in real-time through encoded flow.



## APPROACH

Given *It*, the observation of the robot in the form of monocular RGB image at any time instant *t* and the desired observation *I∗*, our goal is to generate optimal control commands [*vt*,*ωt*] in 6-DoF that minimizes the photometric error between *I∗* and *It*. Instead of directly planning in image observation space, we employ optical flow as an intermediate visual representation for encoding differences in images. We use a pre-trained neural network, Flownet 2 [15] for flow estimation without any fine-tuning. In order to formulate visual servoing as MPC with intermediate flow representations, we define our cost function as mean squared error in flow between any two given images. Then the MPC objective is to generate a set of velocities that minimise the error between the desired flow predicted by the flow network and the generated flow. We use a recurrent neural network (RNN) to generate velocity commands. To train our control network, we employ the predictions from the flow network as targets. Our control network could easily be trained in a supervised manner by minimising our flow loss equation.

*Authors: Pushkal Katara, Y V S Harish, Harit Pandya, Abhinav Gupta*, AadilMehdi Sanchawala*, Gourav Kumar, Brojeshwar Bhowmick, and K. Madhava Krishna*
*Research Center Name: Robotics Research Center*