INTERNATIONAL INSTITUTE OF INFORMATION TECHNOLOGY
HYDERABAD

R&D SHOWCASE 2021
Technology, Social Impact

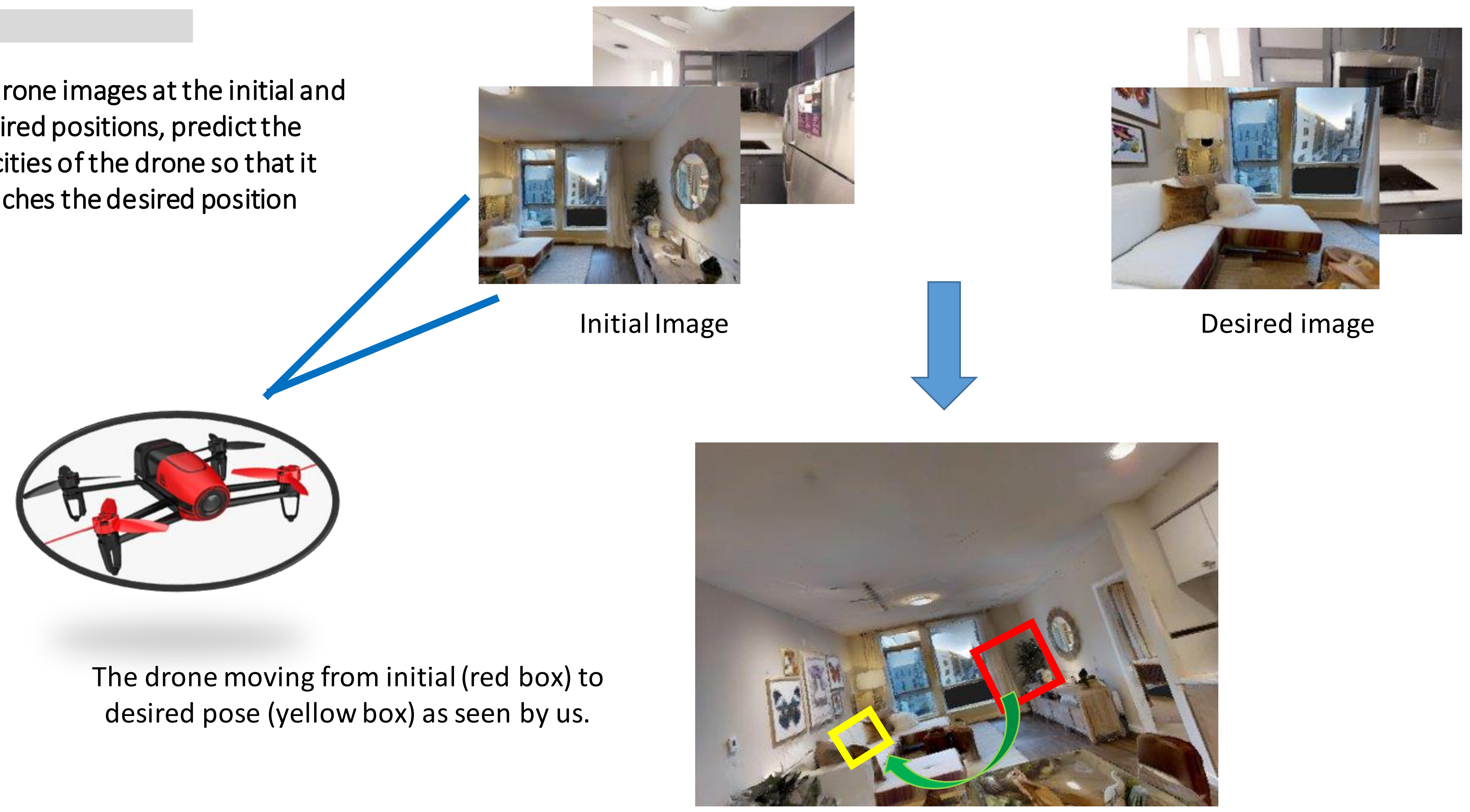
DFVS: Deep Flow Guided Scene Agnostic Image Based Visual Servoing

ABSTRACT

Existing deep learning based visual servoing approaches regress the relative camera pose between a pair of images. Therefore, they require a huge amount of training data and sometimes fine-tuning for adaptation to a novel scene. Furthermore, current approaches do not consider underlying geometry of the scene and rely on direct estimation of camera pose. Thus, inaccuracies in prediction of the camera pose, especially for distant goals, lead to a degradation in the servoing performance. In this paper, we proposed a two-fold solution: (i) We consider optical flow as our visual features, which are predicted using a deep neural network. (ii) These flow features are then systematically integrated with depth estimates provided by another neural network using interaction matrix. We further present an extensive benchmark in a photo-realistic 3D simulation across diverse scenes to study the convergence and generalization of visual servoing approaches. We show convergence for over 3m and 40 degrees while maintaining precise positioning of under 2cm and 1 degree on our challenging benchmark where the existing approaches that are unable to converge for majority of scenarios for over 1.5m and 20 degrees. Furthermore, we also evaluate our approach for a real scenario on an aerial robot. Our approach generalizes to novel scenarios producing precise and robust servoing performance for 6 degrees of freedom positioning tasks with even large camera transformations without any retraining or fine-tuning.

OBJECTIVE

Given drone images at the initial and desired positions, predict the velocities of the drone so that it reaches the desired position



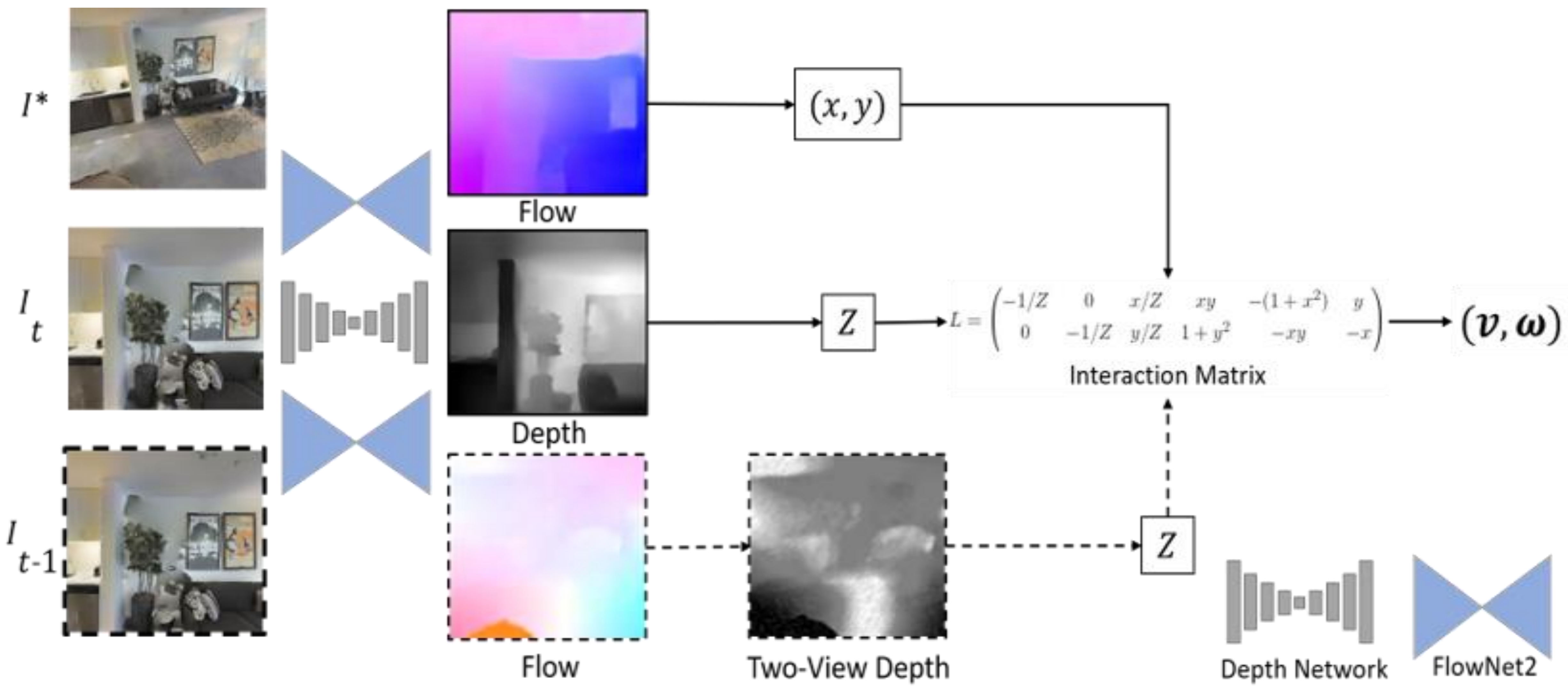
Initial Image

Desired image

The drone moving from initial (red box) to desired pose (yellow box) as seen by us.

METHOD

1. Estimation of the feature error between current and desired pose using Flownet-2.
2. Prediction of the scene's depth at the current pose is done by:
   i. A depth-network that estimates depth from a single view.
   ii. A scaled version of flow is also used as disparity to achieve scene agnostic depth prediction.
3. Generating interaction matrix from the depth and visual features.

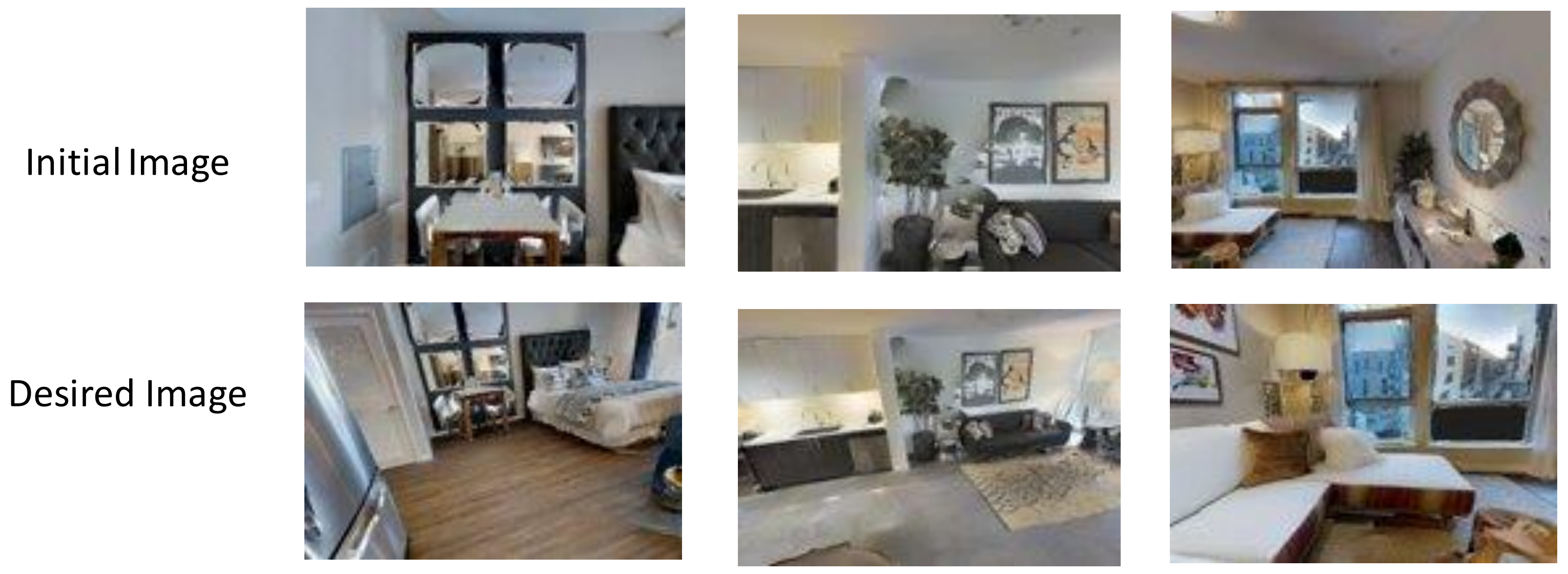


$$L = \begin{pmatrix} -1/Z & 0 & x/Z & xy & -(1+x^2) & y \\ 0 & -1/Z & y/Z & 1+y^2 & -xy & -x \end{pmatrix} \longrightarrow (v, \omega)$$

Baseline Scenes

We classify the complexity of servoing in three categories easy , medium and hard .The categorization was done based on the complexity of the scenes namely the amount of texture present ,the extent of overlap and the rotational , translational complexities between initial and desired image.

We have tested our approach on simulated environments of various complexities and obtained convergence. The images beside depict the hard scenes with a huge amount of rotation[=30°] or translation[>=2m] or both ,and less overlap between images.



Authors: Y V S Harish, Shreya Terupally, Harit Pandya, Ayush Gaud, Sai Shankar, K Madhava Krishna, M, Nomaan Qureshi

ICRA 2020: RRC