CSIAM Trans. Appl. Math.
doi: 10.4208/csiam-am.SO-2025-0064

Vol. 7, No. 3, pp. 600-616
May 2026

Reconstructing Dynamic Scenes with Topologically-Varying Neural Implicit Fields

Kangkan Wang[1], Miao Zhao[1] and Shao-Yuan Li[2,*]

[1] *The Key Laboratory of Intelligent Perception and Systems for High-Dimensional Information of Ministry of Education, School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210014, P.R. China.*
[2] *The MIIT Key Laboratory of Pattern Analysis and Machine Intelligence, College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, P.R. China.*

Received 30 June 2025; Accepted 28 September 2025

Abstract. This paper addresses the challenge of reconstructing dynamic scenes from monocular videos with dynamic neural implicit fields. Previous methods combine a single canonical template with a per-frame deformation field or optimize a hyperspace of templates to cover the variance of the scenes with a large amount of templates. However, single-template based methods cannot recover scenes with topology changes, while hyperspace template approaches may fail to be optimized under large-scale topology changes or motions without any constraints over the high-dimensional space. To address these issues, we propose a topologically-varying neural implicit fields by deforming the sum of multiple canonical templates and recovering large-scale motion and topology changes with a video segmentation strategy. Specifically, multiple templates are defined in canonical space, i.e. a primary template and sparse auxiliary templates. The primary template represents parts that do not change significantly during movement, while the auxiliary templates are combined linearly to represent topological changes. To reconstruct large-scale topologically-varying scenes, a dynamic segmentation strategy is adopted to divide the entire video into multiple segments, and each segment is modeled with its own templates. Compared to existing reconstruction methods for topologically changing objects, our method uses sparse auxiliary templates to form a continuous hyperspace, which is easier to optimize even under large topology changes and motions. Experimental results on various datasets demonstrate that our method outperforms previous works in both geometric reconstruction and novel-view synthesis.

AMS subject classifications: 94A08

Key words: Dynamic reconstruction, topologically-varying neural implicit fields, monocular videos.

*Corresponding author. *Email address:* lisy@nuaa.edu.cn (S. Li)

1 Introduction

Dynamic scene reconstruction from monocular videos is an important problem in the fields of computer vision and graphics, and it has many promising applications such as virtual reality, augmented reality, and video games. Given a monocular video of moving objects, the goal is to reconstruct sequential 3D models of the dynamic objects with consistent geometry and appearance over time. This is an extremely challenging task due to arbitrary motion of the objects, serious depth ambiguity under monocular setting and topological changes in both geometry and appearance (e.g. the hand appears when the glove is taken off).

Previous reconstruction systems rely on expensive multi-camera studios [4, 11, 20, 27, 28] that prevent their daily applications or require depth sensors [8, 14, 36, 37, 39] for high-quality 3D reconstruction which cannot be applied when 3D data is unavailable. Recent methods [17, 21, 22] from monocular or multi-view videos use an implicit 3D representation to represent a canonical template and generate dynamic reconstructions by deforming the implicit representation to video frames with a deformation fields. However, a single template used in these methods is not sufficient to represent all possible varying states of moving objects, and the linear nature of the deformation fields requires deformations to be uniform which cannot account for topological changes or large motions. Recently, HyperNeRF [18] extends the 3D template to a high-dimensional template space, and recovers topologically-changing surfaces by learning hyper-space templates together with pre-frame deformation fields. However, the hyperspace is infinitely large and lacks constraints on the optimization of templates. For videos with large motion and topological changes, the varying space of templates increases significantly, making it difficult to optimize without any constraints over the hyper space.

In this paper, we propose a novel method for dynamic scene reconstruction with topologically-varying neural implicit fields. Specifically, based on high-quality geometry representation of the neural implicit surface (NeuS) [29], we learn multiple templates represented by NeuS in canonical space and the linear composition parameters of templates as well as deformation fields of each video frame by matching the rendered color with the observed color. To effectively handle topological changes, we decompose the template into a primary template to represent the invariant or small changing parts of the object, and a linear combination of several sparse auxiliary templates to represent the topology-changing parts. To reconstruct objects with large motion and large-scale topology changes, we dynamically segment the input video based on per-frame color errors and define the corresponding template set for each new segment (each set consisting of a primary template and several sparse auxiliary templates) to reconstruct the scene in each segment. During the training on an input frame, we first select the corresponding set of templates according to the segment index, and then obtain the final geometry and color by linearly compositing the values fetched on the primary and sparse auxiliary templates. Experimental results on our own various datasets demonstrate that our method can reconstruct high-quality geometry and appearance from monocular