JULY 19–22
VANCOUVER, BC
2026 JOINT AAPM | COMP MEETING

Transformer-based Multi-channel Target Decomposition for Robust Markerless Lung Tumor Tracking

Jiening Zhu, Bohong Huang, Pengpeng Zhang, Qiyong Fan, Weixing Cai, Hai Pham, Laura Cervino, Jean M. Moran, Tianfang Li, Xiang Li, Yabo Fu

Department of Medical Physics, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA

ABSTRACT

Purpose: This study aims to develop a deep learning–based framework for markerless lung tumor tracking that improves tracking accuracy, robustness, and computational efficiency to enable seamless clinical integration.
Methods: We developed a transformer-based framework that maps raw kV projection images into multiple decomposed target images (DTIs), where each DTI encodes a distinct spatial context relative to the tumor location. DTIs are synthesized by digitally reconstructing thin-volume slabs with varying thicknesses and spatial offsets relative to the target. The model was trained using digitally reconstructed radiography (DRR)–DTI image pairs generated from simulation CT data and tested on real kV images acquired using a Varian TrueBeam on-board imager (OBI). Multiple synthetic DTIs were generated for each kV image using the trained model. Parallel template matching is performed on these synthetic DTIs to generate multiple candidate tumor positions. These candidates are subsequently fused using an adaptive extended Kalman filter (AEKF), which integrates multiple current measurements with one-step motion history to estimate the most probable tumor location while providing real-time localization uncertainty. Our method was evaluated using a chest motion phantom with known ground-truth tumor motion and on clinical patient data with implanted Calypso beacons adjacent to the tumor serving as ground truth.
Results: Our method achieved a maximum tracking error of 1.14 mm for the phantom study. For patients, our method achieved a 94.5% tracking success rate across 4,312 images from nine patients. Tracking success was defined as a localization error of less than 2.0 mm in the superior–inferior (SI) direction.
Conclusion: This study demonstrates that transformer-based spatial decomposition substantially enhances the localization of low-contrast lung tumors in kV projection images. The achieved high accuracy, robustness, and inherent uncertainty estimation indicate that the proposed framework is a strong candidate for real-time intrafraction motion management in high-precision radiotherapy.

Memorial Sloan Kettering Cancer Center

CONTACT

Jiening Zhu
Department of Medical Physics
Memorial Sloan Kettering Cancer Center,
New York, NY
Email: zhuj3@mskcc.org

INTRODUCTION

Background

Markerless lung tumor tracking using kV projection imaging can improve radiotherapy precision without implanted fiducial markers. However, clinical implementation remains challenging due to:

- Low soft-tissue contrast in kV images.
- Tumor occlusion by ribs and other anatomy.
- Single-output deep learning models that are vulnerable to ambiguous images.
- Limited estimation of tracking confidence and uncertainty.

Our Approach

We propose a transformer-based multi-channel target decomposition framework for robust markerless tumor tracking:

- Generate multiple depth-aware target images from a single kV projection.
- Perform parallel template matching to produce multiple localization candidates.
- Fuse candidates with an Adaptive Extended Kalman Filter (AEKF) for robust tracking and real-time uncertainty estimation.
- Train on DRR–target image pairs for efficient clinical deployment.

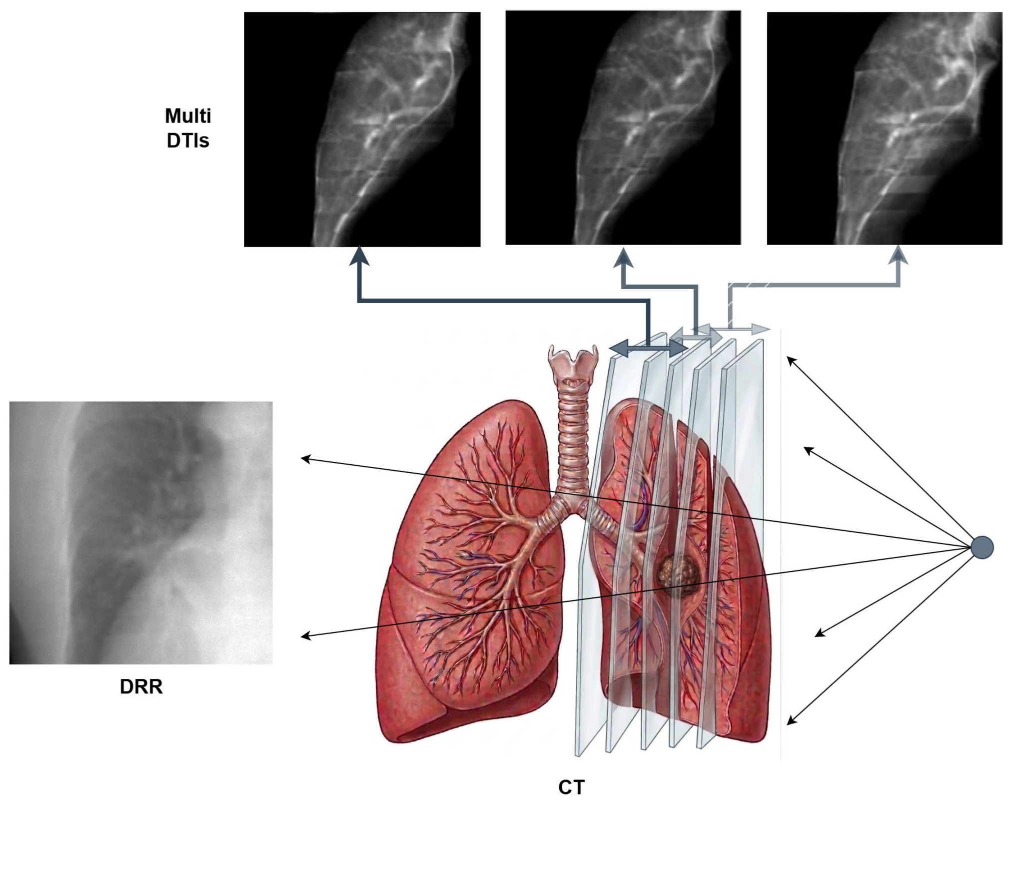

Figure 1. Multi-DTI generation — one kV/DRR projection is decomposed into multiple depth-resolved target images (DTIs) reconstructed from thin CT slabs of varying thickness and offset.

INNOVATIONS

- Multi-Channel DPT uses global attention to resolve the ill-posed 2D→2.5D problem and localize low-contrast soft-tissue targets.
- Adaptive Extended Kalman Filter fuses parallel multi-context tracks, mitigating discontinuities and giving real-time uncertainty.
- On-the-fly, multi-GPU training pipeline removes disk storage and enables richer augmentation for better generalization.

METHODS

Multi-View Tumor Localization and Fusion

The synthesized DTIs are independently analyzed to estimate tumor position, after which the multiple localization results are combined into a single robust tracking solution.

- Perform parallel template matching on each synthesized DTI to generate multiple candidate tumor locations.
- Fuse complementary measurements using an Adaptive Extended Kalman Filter (AEKF), incorporating both current observations and temporal motion history.
- Estimate real-time localization uncertainty by evaluating the agreement among multiple tracking hypotheses.
- Validate the framework using a motion phantom with known ground truth and clinical kV projection images with Calypso beacon localization as the reference standard.

Multi-View Tumor Localization and Fusion

The synthesized DTIs are independently analyzed to estimate tumor position, after which the multiple localization results are combined into a single robust tracking solution.

- Perform parallel template matching on each synthesized DTI to generate multiple candidate tumor locations.
- Fuse complementary measurements using an Adaptive Extended Kalman Filter (AEKF), incorporating both current observations and temporal motion history.
- Estimate real-time localization uncertainty by evaluating the agreement among multiple tracking hypotheses.
- Validate the framework using a motion phantom with known ground truth and clinical kV projection images with Calypso beacon localization as the reference standard.

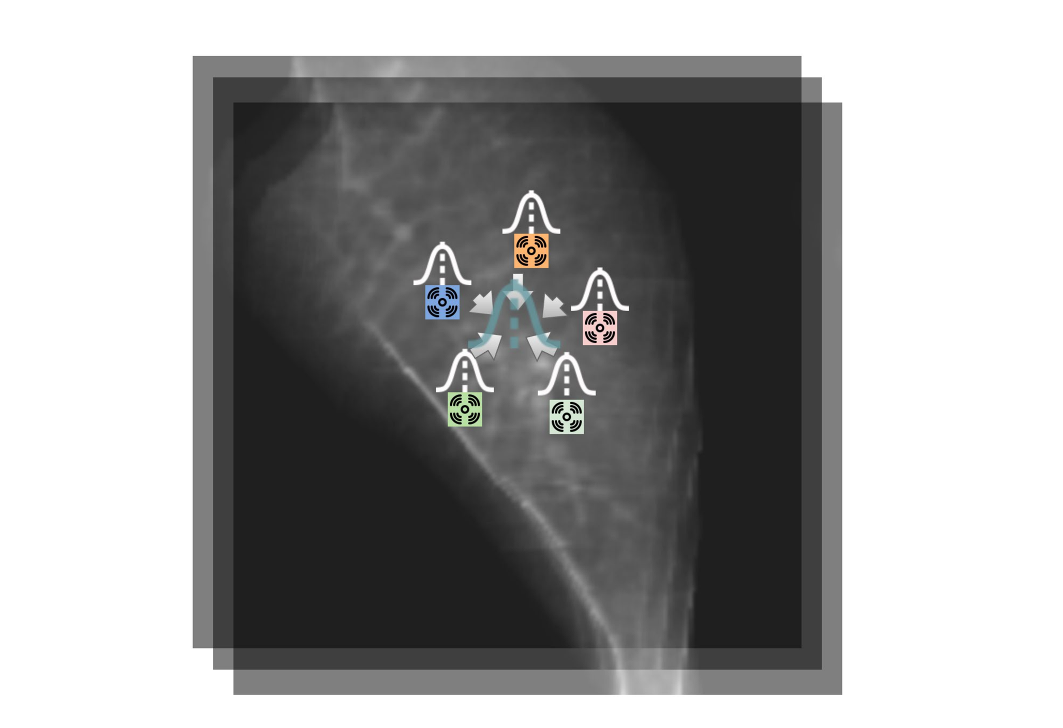

Figure 2. Parallel tracking and fusion — each DTI is tracked by template matching; candidates are fused by an adaptive EKF into one location estimate with uncertainty.

RESULTS

The proposed framework was evaluated using both a dynamic breathing motion phantom with known ground-truth tumor motion and clinical kV projection images from patients with implanted Calypso beacons serving as the reference standard. The method demonstrated high accuracy and robustness across both datasets, achieving a **94.5% tracking success rate** in clinical evaluation while maintaining **sub-millimeter** localization accuracy in phantom experiments. Compared with previous deep learning–based approaches, the transformer-based multi-channel framework consistently improved tumor localization, particularly in challenging cases with low soft-tissue contrast and severe anatomical overlap. The Adaptive Extended Kalman Filter further enhanced tracking stability and provided real-time localization uncertainty. Detailed quantitative results are summarized in **Table 1**.

Table 1. Tracking accuracy on phantom and patient data

Study	Condition	MAE (mm)
Phantom	DIBH	0.202
Phantom	Free-breath	0.232
Phantom	Sinusoidal	0.255
Patient	Free-breath	0.7

94.5% patient success rate (SI error < 2 mm)

1.14 mm max phantom tracking error

4,312 kV images 9 patients

CONCLUSIONS & DISCUSSIONS

This work presents a transformer-based multi-channel target decomposition framework for robust markerless lung tumor tracking from kV projection images. By combining complementary depth-aware target representations with adaptive multi-hypothesis fusion, the proposed method achieves accurate and reliable tumor localization while providing real-time uncertainty estimation. The strong performance on both phantom and clinical datasets demonstrates its potential for clinical translation in image-guided radiotherapy.

Future work will focus on **real-time patient-specific adaptation**, including rapid **per-fraction fine-tuning** using pretreatment imaging and **progressive online fine-tuning** during treatment to continuously adapt to anatomical and respiratory motion changes. These developments are expected to further improve tracking robustness and support routine clinical implementation.