WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Video Editing Framework Using Object Trajectories in 3D Space

Miss. Kamble Ragini Vinayak[1], Dr. Sushilkumar N. Holambe[2]

[1]Student, M.Tech Dept of Computer science and Engineering , TPCTs college of engineering Dharashiv.

[2](Associate Prof. & Dean R&D), M.Tech Dept of Computer science and Engineering , TPCTs college of engineering Dharashiv.

Abstract:

Conventional video editing platforms usually treat videos as a sequence of frames aligned on a timeline, which makes object-level modifications tedious and inefficient. In this paper, we present an object-centric framework for manipulating long-shot videos by introducing operations at three semantic levels: background mosaics, object trajectories, and camera movements. The static scene background is extracted independently, while dynamic objects are represented as 3D space–time trajectories, forming the primary units for interaction. Users can perform temporal adjustments and spatial edits through direct trajectory manipulation. In addition, the camera is modeled as a movable and scalable viewport, allowing virtual pan, tilt, and zoom operations by defining new camera paths. Our experiments demonstrate that this approach simplifies complex video editing tasks, reduces manual effort, and provides a more intuitive and flexible editing experience.

Keywords: Interactive video editing, Object-centric framework, 3D trajectories, Camera motion modeling, Background mosaics, Temporal and spatial manipulation, Video composition.

Introduction

Conventional video editing tools typically operate on frame-by-frame sequences arranged along a timeline. While adequate for basic cutting, sequencing, and transitions, this approach becomes cumbersome when precise, object-level manipulation is required. Tasks such as repositioning a moving subject, synchronizing multiple objects, or modifying camera viewpoints often involve painstaking frame-by-frame adjustments, increasing editing time and requiring technical expertise. Recent advances in computer vision—such as object detection, motion segmentation, and background modeling—allow extraction of higher-level semantic information from video content. Leveraging these techniques, we propose a framework that emphasizes object-centric trajectories and semantic camera modeling, enabling direct and intuitive interaction for complex video edits.

Contributions of this work:

1. Unified representation separating static backgrounds, moving objects, and camera motion.
2. Use of 3D space–time trajectories as intuitive handles for object manipulation.
3. Camera-as-aperture modeling enabling virtual pan, tilt, and zoom.
4. Demonstration of real-world editing scenarios with simplified workflows and reduced manual effort.

Hardware and Software Requirements

Hardware Requirements

- Processor: Intel i5/i7 or equivalent
- RAM: 16 GB or more
- GPU: NVIDIA GTX 1050 or higher for trajectory computation
- Storage: Minimum 500 GB

Software Requirements

- Operating System: Windows 10 / Linux
- Video Editing Framework: OpenCV / FFmpeg
- Programming Language: Python / C++
- IDE: Visual Studio / PyCharm
- Additional Tools: Media player for previewing output, optional LaTeX for documentation

Proposed Method

4.1 Background Estimation

Static regions are extracted across frames to create a background mosaic. This ensures clear separation between moving objects and the scene, simplifying editing operations.

4.2 Object Trajectory Representation

Dynamic objects are tracked over time to form 3D space–time trajectories. These trajectories serve as intuitive handles for temporal and spatial editing, allowing speed adjustment, path modification, and synchronization with other objects.

4.3 Camera Motion Modeling

The camera is treated as a movable and scalable viewport. Users can simulate pan, tilt, and zoom operations by defining new camera trajectories, enabling flexible scene composition without additional recording.

4.4 Interaction Operators

- Temporal shifting and path redirection of objects
- Camera path adjustment for smooth virtual pan, tilt, and zoom
- Integration of multiple object and camera edits into the final video

Future Work

1. **Real-time Processing:**
   Currently, the framework operates offline. In the future, trajectory extraction and camera modeling could be implemented in real-time, enabling applications in live video editing or streaming.
2. **AI-Assisted Editing:**
   Integrating AI and deep learning for object recognition, scene understanding, and motion prediction could allow the system to automatically suggest edits, further reducing manual effort for editors.
3. **Support for 360° and VR Videos:**
   The current framework is designed for standard 2D videos. Extending it to support 360° and VR content would enable immersive video editing capabilities.
4. **Enhanced Multi-Object Interaction:**
   Improvements in multi-object trajectory synchronization, collision detection, and automated path planning could facilitate more complex scene manipulations in an intuitive manner.
5. **Advanced Camera Effects:**
   Additional camera path modeling could support advanced effects such as smooth dolly, crane, and orbital camera movements for cinematic compositions.
6. **Improved User Interface:**
   Incorporating drag-and-drop trajectory editing, gesture-based controls, and other interactive UI enhancements could improve usability and make the editing process more intuitive.

Results and Discussion

The framework was tested on various video types, including sports sequences, surveillance footage, and cinematic scenes:

- Editing time was reduced compared to conventional frame-by-frame approaches.
- Object movements and camera effects were smooth and visually consistent.
- Users found the interface intuitive, enabling creative and flexible edits.

Conclusion

We presented an interactive object-centric video editing framework using 3D trajectories and background mosaics. The system transforms manual frame-by-frame edits into intuitive curve-based interactions, enabling both object and camera manipulations efficiently. Future extensions include real-time AI-assisted trajectory prediction and support for 360° and VR video editing.

References:

1. R. Szeliski, "Image Alignment and Stitching: A Tutorial," *Foundations and Trends in Computer Graphics and Vision*, vol. 2, no. 1, pp. 1–104, 2006.
2. B. K. Horn and B. G. Schunck, "Determining Optical Flow," *Artificial Intelligence*, vol. 17, no. 1–3, pp. 185–203, 1981.
3. Z. Wu, T. Y. Lim, and M. Brown, "Video Object Manipulation with Trajectory Editing," *IEEE Trans. Vis. Comput. Graph.*, vol. 24, no. 4, pp. 1457–1468, 2018.

4. A. Rav-Acha, Y. Pritch, and S. Peleg, "Making a Long Video Short: Dynamic Video Synopsis," CVPR, 2006.
5. C. Liu, J. Yuen, and A. Torralba, "SIFT Flow: Dense Correspondence across Scenes and Its Applications," *IEEE Trans. PAMI*, vol. 33, no. 5, pp. 978–994, 2011.
6. A. Agarwala et al., "Interactive Digital Photomontage," *ACM Trans. Graph.*, vol. 23, no. 3, pp. 294–302, 2004.
7. Y. Boykov and V. Kolmogorov, "An Experimental Comparison of Min-Cut/Max-Flow Algorithms for Energy Minimization in Vision," *IEEE Trans. PAMI*, vol. 26, no. 9, pp. 1124–1137, 2004.
8. C. Tomasi and T. Kanade, "Detection and Tracking of Point Features," *Int. J. Comput. Vis.*, vol. 9, no. 2, pp. 137–154, 1991.