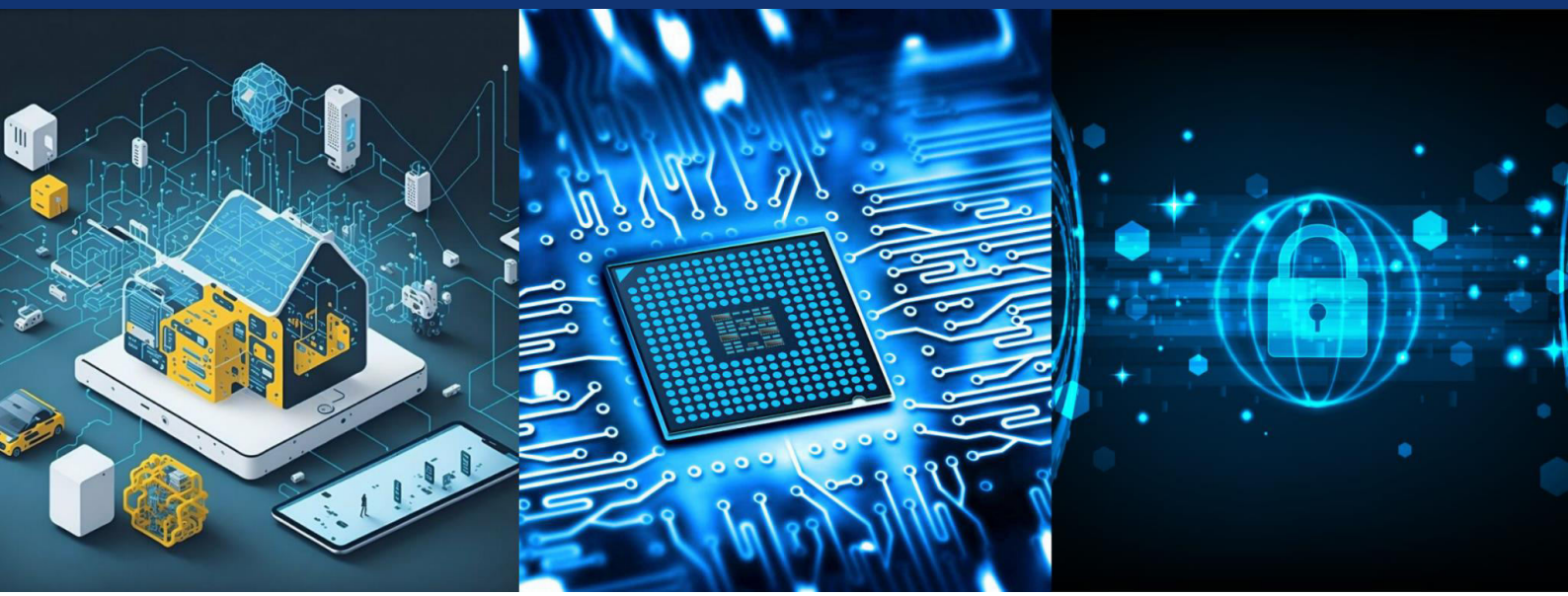
IJIRCCE

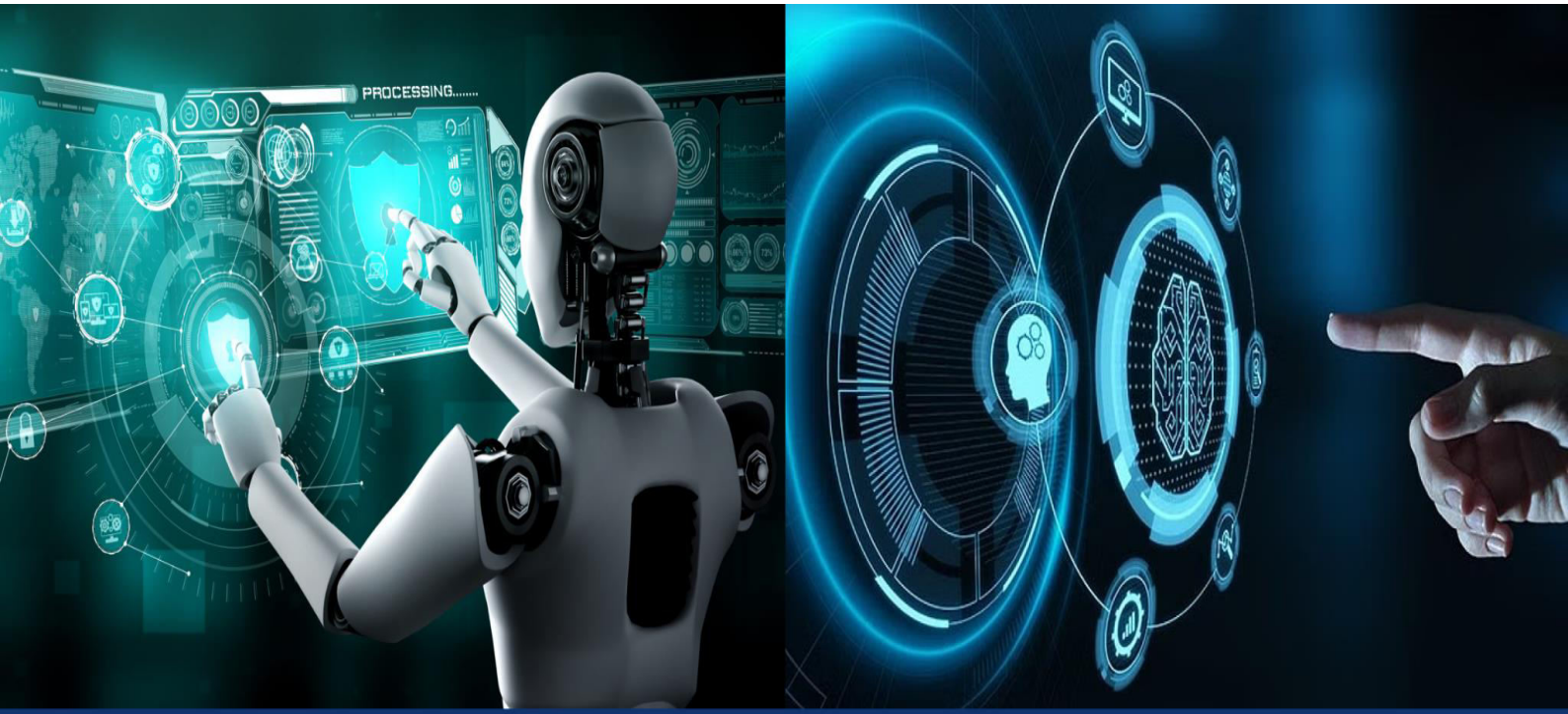
ISSN(O): 2320-9801

ISSN(P): 2320-9798



International Journal of Innovative Research in Computer and Communication Engineering

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.771

Volume 13, Issue 4, April 2025

© 2025 IJIRCCE | Volume 13, Issue 4, April 2025| DOI: 10.15680/IJIRCCE.2025.1304074

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Deep Convolutional Networks using Gans for Rendering of Human Faces

Dr. B. Jogeswara Rao, T. Naveena, K. Navya Sree, P. Pranith, J. Nagendra

Associate Professor, Department of CSE, Malla Reddy University, Hyderabad, India

B. Tech, School of Engineering, Malla Reddy University, Hyderabad, India

ABSTRACT: This paper presents a novel deep learning system for animating still images using motion extracted from driving videos in an unsupervised manner. Traditional approaches relying on key point detection face limitations with articulated objects and identity preservation. To address this, we introduce a Principal Component Analysis (PCA)-based region motion estimation method that generates affine transformations on motion heatmaps, enabling more semantically meaningful motion representations. Additionally, our framework employs a disentanglement strategy to separate shape and pose, preserving the identity of the subject image, and a dedicated background motion estimator to reduce visual artifacts. Evaluation conducted on benchmark datasets such as TaiChiHD and VoxCeleb demonstrates superior performance in motion realism, identity retention, and animation coherence compared to state-of-the-art models like FOMM. This research provides a scalable and generalizable framework suitable for applications in AI avatars, telepresence, media generation, and education.

KEYWORDS: Cyberbullying, Social Media, BERT, NLP, Semi-supervised learning , Twitter API.

I. INTRODUCTION

Animating still images using motion transfer is an emerging field with impactful applications in virtual avatars, digital content creation, and synthetic media. Most conventional approaches are based on either manual rigging or keypoint estimation, which are prone to identity leakage and ineffective motion transfer, especially for articulated figures like human bodies and faces.

The concept of transferring motion from a video to a still image involves capturing temporal information from the video and applying it spatially onto a static frame. Earlier solutions often focused on extracting human pose using keypoints or predefined motion graphs. However, these methods fail in representing complex and non-rigid motion patterns. Our work is inspired by the advancements in unsupervised deep learning techniques that do not require labeled datasets and instead use inherent motion patterns within video sequences.

The core methodology includes these three critical contributions: (1) PCA-based motion of estimation to extract affine transformations from the heatmaps, (2) and explicit background motion modeling to isolate the foreground object, and also (3) animation via disentanglement to preserve object shape while applying motion from a driving video. The system is implemented using deep learning frameworks with components such as region predictors, background motion estimators, and dense optical flow predictors. The training is conducted in a self-supervised manner using reconstruction and equivariance losses.

This paper presents a novel approach based on the research work by Siarohin et al. titled "Motion Representations for Articulated Animation." We enhance image animation by adopting PCA-based region motion modeling, background motion isolation, and disentangled feature representations. The system is trained using unsupervised learning on video datasets, removing the need for labeled data.

OBJECTIVES:

- This project focuses on enhancing motion estimation by replacing traditional keypoint-based techniques with a region-based Principal Component Analysis (PCA) approach. Unlike keypoint detection, which often lacks semantic consistency, PCA-based motion estimation captures movement across meaningful regions, resulting in more natural and realistic animations.

© 2025 IJIRCCE | Volume 13, Issue 4, April 2025| DOI: 10.15680/IJIRCCE.2025.1304074

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

- A crucial goal is to separate shape and pose during the animation process. By processing identity (shape) and motion (pose) independently, the system preserves the original structure of the source image, preventing distortions that can occur when motion is directly transferred.
- To improve animation accuracy, the system includes an independent background motion estimator. This feature effectively separates background movement from foreground motion, minimizing unwanted motion artifacts and ensuring precise animation of the subject.
- The project also adopts self-supervised learning techniques, removing the need for manually labeled datasets. This approach enhances adaptability and scalability, allowing the model to work effectively with various inputs without requiring extensive human annotations.
- Another objective is to achieve superior animation quality compared to existing methods like the First Order Motion Model (FOMM). By refining motion synthesis techniques, the system generates smoother transitions and more visually appealing animations with minimal distortions.

Finally, this project explores practical applications in areas such as virtual avatars, educational tools, and personalized media creation. These implementations highlight the system's potential in fields like entertainment, e-learning, and digital content development, demonstrating its versatility and effectiveness in real-world scenarios.

II. PROBLEM STATEMENT

Animating static images using deep learning has become a fascinating research domain, with significant implications for virtual reality, video editing, and synthetic media. While models like the First Order Motion Model (FOMM) have demonstrated notable progress, animating complex, articulated objects—such as human faces and bodies—remains a formidable challenge. The intricacies of non-rigid motion often pose difficulties for existing approaches, limiting their ability to produce natural and coherent animations.

Traditional methods primarily rely on keypoint detection to track the movement. However, this approach has also have fundamental shortcomings. Keypoints typically capture motion in a way that emphasizes edges and arbitrary landmarks, often failing to represent the actual structure and semantics of an object. Additionally, these models are highly susceptible to background variations and camera shifts, which can introduce unwanted distortions. Another common issue is the unintended transfer of shape attributes from the driving video to the source image, resulting in identity leakage and reduced visual fidelity. These limitations will become particularly pronounced when dealing with articulated figures, where accurately modeling the movement of distinct body parts—such as arms, legs, and facial features—is essential for realism.

Furthermore, background motion remains an unsolved problem in many existing frameworks. Real-world videos often involve the dynamic environments with camera movements and also changing elements, yet many models fail to separate foreground motion from background changes. This oversight leads to motion artifacts, ultimately diminishing the quality of the animation.

Existing image animation methods predominantly depend on keypoints or skeletal representations to capture and transfer motion. While these methods can animate objects to some extent, they often:

- Fail to accurately represent articulated and non-rigid motion such as those in human limbs and facial expressions.
- Introduce identity distortion by transferring not just pose but also structural attributes from the driving video.
- Suffer from motion artifacts due to camera movement or complex background dynamics.
- Require labeled datasets, which are often costly and difficult to obtain.

Therefore, the challenge lies in creating an unsupervised model that:

- Accurately captures semantically meaningful motion without relying on labeled keypoints.
- Preserves the identity and shape of the source image.
- Separates and models background motion to reduce noise in animation.
- Generalizes across diverse datasets involving different human activities and backgrounds.

III. LITERATURE SURVEY

To understand the evolution of motion animation models, we reviewed several key papers:

First Order Motion Model (FOMM) – Siarohin et al. (2019): Introduced a novel unsupervised method using keypoints

and motion fields to animate images. While foundational, the model faced limitations in identity retention and struggled with complex motion.

Motion Representations for Articulated Animation – Siarohin et al. (2021): The basis of our current project, this paper replaces keypoints with PCA-based region motion, adds background motion modeling, and proposes disentanglement of shape and pose.

GANimation – Pumarola et al. (2019): Focused on facial expression animation using action units. Provided fine-grained control over expressions but was limited to faces.

Talking Head Models – Zakharov et al. (2019): Introduced few-shot learning to generate talking-head animations from limited data. Emphasized identity preservation and GAN-based adversarial training.

X2Face – Wiles et al. (2018): Used appearance and pose encoders to transfer motion. Provided good generalization for faces but lacked semantic segmentation.

Deformable GANs – Siarohin et al. (2018): Enabled pose-based animation through skip connections and deformation fields.

Key Insights and Gaps:

- Keypoint-based models often fail with non-rigid articulated motion.
- Identity leakage is a major challenge.
- Limited handling of background motion.
- Need for unsupervised, generalizable models.

Our project builds upon these findings to deliver a more robust solution.

IV. PROPOSED SYSTEM

The proposed system introduces a region-based motion representation model designed to animate articulated objects from still images. This method enhances motion accuracy, preserves identity, and improves animation stability. The system builds upon three key innovations inspired by Siarohin et al.

1. PCA-Based Motion Estimation

Traditional keypoint regression methods often struggle with maintaining motion stability and accuracy. To address this, the system leverages Principal Component Analysis (PCA) on region heatmaps instead of detecting discrete keypoints. PCA enables the computation of affine transformations, ensuring that motion representations are more stable and semantically meaningful. This approach leads to smoother and more realistic animations by capturing motion across meaningful regions rather than arbitrary landmarks.

2. Background Motion Modeling

A major challenge in motion transfer is distinguishing between object movement and background shifts. The system includes a dedicated background motion estimator, which isolates environmental and camera motion from the motion of the animated object. By preventing unnecessary background distortions, this feature significantly enhances animation quality and reduces unwanted artifacts, especially in dynamic scenes with camera movements.

3. Disentangled Animation Framework

To maintain the structural integrity of the source image, the system employs a disentangled animation framework with separate encoders for shape and pose. The shape encoder ensures the source identity remains unchanged, while the pose encoder extracts and applies motion information from the driving video. This separation prevents structural distortions and ensures high-fidelity motion transfer.

4. Deep Learning Pipeline and Training

The system integrates these components into a deep learning pipeline, which consists of region predictors, optical flow generators, and image synthesis modules. Training is conducted using self-supervised learning, incorporating reconstruction and equivariance losses to improve model generalization. The architecture is also designed to scale efficiently with varying numbers of motion regions, making it adaptable to different datasets while maintaining high

© 2025 IJIRCCE | Volume 13, Issue 4, April 2025| DOI: 10.15680/IJIRCCE.2025.1304074

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

animation quality.

V. SYSTEM DESIGN

System design plays a critical role in software development, bridging the gap between conceptual ideas and practical implementation. In this project, which focuses on animating static images using motion patterns extracted from a driving video, the system design phase establishes the structure and interaction of key components such as data processing, motion representation, animation synthesis, and output generation.

This section details the system design based on the methodology outlined in the research paper *"Motion Representations for Articulated Animation"* by Siarohin et al. It covers the system's perspective, architectural framework, data flow, operational sequence, use cases, and flowcharts for better visualization of the overall logic.

The system is structured as a modular deep learning application that enables users to upload images and videos for animation generation. It functions as a standalone tool but can be integrated into broader platforms like multimedia editing software or avatar creation systems. The system processes external datasets and can leverage cloud-based environments like Google Colab for both training and inference.

The primary target users include students, researchers, and developers seeking an efficient image animation solution. Additionally, an administrative interface may be available for developers, allowing model retraining and management.
The architecture follows a modular design:

- **Input Layer** The input layer receives a static image and a driving video, serving as the foundation for animation. It ensures compatibility with subsequent processing steps, enabling seamless extraction of motion features for synthesis.
- **Preprocessing Module**This module standardizes inputs by resizing images, normalizing pixel values, and extracting frames from the driving video. These preprocessing steps enhance data consistency, ensuring smoother motion estimation and reducing discrepancies between source and driving frames.
- **Region Detection** The system generates unsupervised heatmaps to identify key motion regions within the image. By detecting movement-sensitive areas, this step ensures accurate tracking of articulated parts, improving animation realism and motion coherence.
- **PCA Module** A Principal Component Analysis (PCA)-based module computes affine transformation matrices for detected motion regions. This step models local motion patterns, capturing subtle shifts in shape and movement while preserving structural integrity across frames.
- **Encoders** Two separate encoders process shape and pose information, preventing identity distortion. The shape encoder retains the source image's structure, while the pose encoder extracts motion patterns from the driving video for precise animation transfer.
- **Optical Flow Estimator** This component generates dense motion maps by estimating pixel-wise displacement between frames. By refining motion accuracy, it ensures fluid transitions, eliminating jittery movement and preserving natural motion dynamics in the animated output.
- **Decoder** The decoder synthesizes new frames by warping source image features according to the computed motion maps. It reconstructs realistic transformations, ensuring the animated sequence maintains fidelity to both the original appearance and intended motion.
- **Output Module** The output module assembles the generated frames into a cohesive animated video. It sequences frames correctly, renders the final output, and provides a downloadable file, ensuring accessibility for users across various applications.

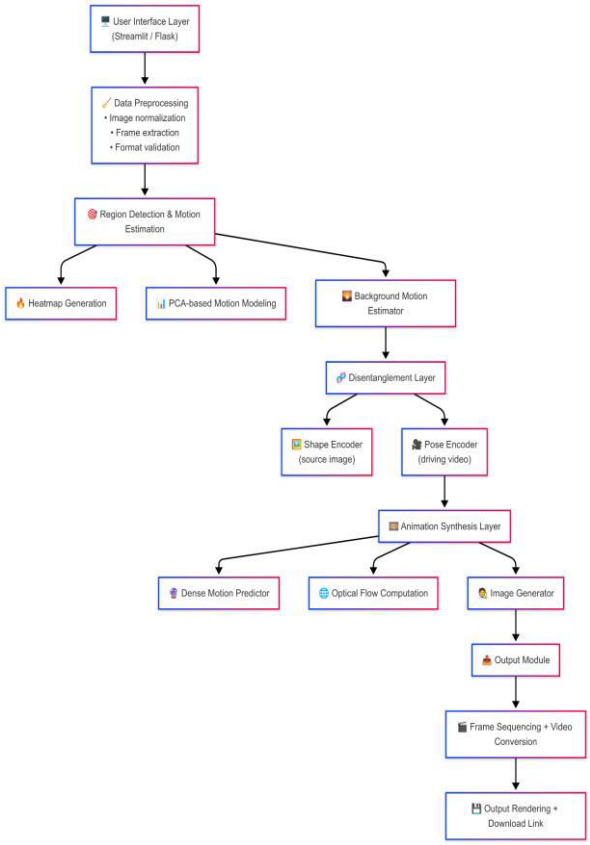
This modular architecture ensures maintainability, flexibility, and scalability.

© 2025 IJIRCCE | Volume 13, Issue 4, April 2025| DOI: 10.15680/IJIRCCE.2025.1304074

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



VI. IMPLEMENTATION

- **Platform**: Google Colab
- **Languages/Tools**: Python, PyTorch, OpenCV, NumPy, ffmpeg
- **Training**: Conducted on video datasets using self-supervised losses (reconstruction, perceptual, and equivariance losses).
- **Preprocessing**: Automated frame extraction and normalization.
- **Animation Generation**: Intermediate frames synthesized using warped features and optical flow.
- **Deployment**: Prototyped with Colab UI; can be extended to Streamlit for production.

Flowchart

The flowchart below outlines the step-by-step process involved in the image animation system. It provides a visual representation of how data and operations flow from the moment the user initiates the session to the generation of the final animated video:

- Start: The system is initialized, and the session begins.
- Upload Source Image: The user provides a still image that serves as the base for animation.
- Upload Driving Video: The user uploads a video that provides motion cues.
- Preprocess Inputs: The system processes both inputs by resizing, normalizing, and extracting frames.
- Generate Heatmaps from Video: The model extracts region-specific motion features using unsupervised learning.
- Apply PCA for Motion Extraction: PCA is applied to reduce dimensionality and generate affine transformations.
- Disentangle Shape & Pose: The shape from the source image and pose from the driving video are separated to maintain identity.
- Predict Dense Motion / Optical Flow: Dense motion fields are computed to represent pixel-wise movements.
- Generate Animated Frames: The static image is animated using the motion data, producing sequential frames.

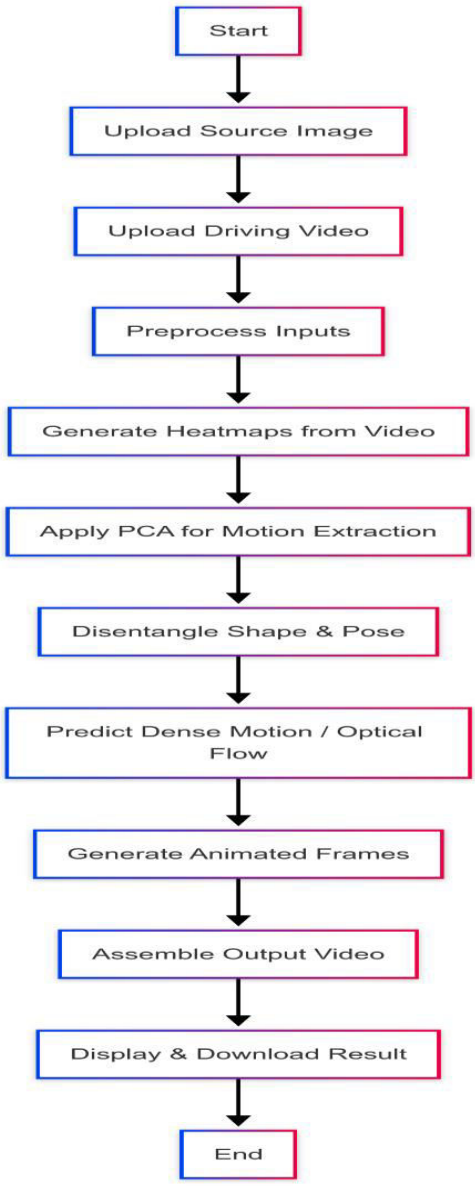
© 2025 IJIRCCE | Volume 13, Issue 4, April 2025| DOI: 10.15680/IJIRCCE.2025.1304074

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

- Assemble Output Video: The frames are compiled into a video format (e.g., MP4).
- Display & Download Result: The animated video is displayed for preview and can be downloaded.
- End: The session concludes.



VII. RESULTS & TESTING

- **Datasets Used**: TaiChiHD, VoxCeleb, TED Talks.
- **Metrics**:
  - Perceptual Similarity Index
  - Structural Similarity Index (SSIM)
  - User Study Ratings
- **Testing Strategy**:
  Manual testing for functional correctness and quality.
  Stress tests with high-resolution inputs.

© 2025 IJIRCCE | Volume 13, Issue 4, April 2025| DOI: 10.15680/IJIRCCE.2025.1304074

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Validation against baseline (FOMM).

Test Case	Description	Result
TC01	Valid inputs	Pass
TC02	Unsupported file	Pass
TC03	Missing video	Pass
TC04	Full pipeline	Pass
TC05	Identity preservation	Pass

Use Case Scenarios

Use Case 1: Image Animation for Personal Avatar

Actor: User Description: A user wants to animate their photo using a reference video to create a speaking avatar. Steps:

- Upload a selfie image.
- Upload a video of a speaking face.
- Generate animation.
- Preview and download the result.

Types of Tests Conducted

Input Validation Tests

- Uploaded images and videos were tested for supported formats (JPG, PNG, MP4).
- Images with very high or low resolution were tested to evaluate how well the system resized and normalized data.

Functionality Tests

- Preprocessing logic was verified by visually inspecting normalized frames.
- Motion estimation module was evaluated by checking if PCA components changed based on motion content.
- Frame generation accuracy was tested by inspecting the realism of animations.

Output Quality Tests

- Animated videos were visually inspected for smoothness, frame coherence, and identity preservation.
- Compared output results across different source images and driving videos to assess model generalization.

Stress and Limit Tests

- Long videos and high-resolution images were used to test memory usage and Colab runtime limits.
- The system's behavior under slow internet conditions was observed.

Negative Testing

- Uploaded unsupported file formats to observe system behavior.
- Provided only one input (either image or video) to test input validation handling.

Test Results and Observations

Successes:

- The system consistently produced coherent and smooth animations when standard resolution inputs were provided.
- Shape-pose disentanglement worked well in preserving the identity of the source image.
- PCA-based motion estimation yielded better control of regional movement than keypoint methods.

Minor Issues:

- Slight flickering occurred in animations when the driving video had extreme movements.
- Colab occasionally timed out for longer animations, requiring re-execution.

Failures:

- Uploading corrupted or incomplete files led to execution errors.
- System failed to animate properly if the driving video lacked significant motion.

VIII. DISCUSSION

Since the system was developed and executed within Google Colab, testing was performed manually, focusing on functional accuracy, output quality, and robustness against diverse inputs. The testing framework was structured into multiple stages to ensure comprehensive validation:

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

- Module-Level Testing: Each system component—such as preprocessing, motion estimation, and frame generation—was individually tested to verify its correctness and performance.
- Integration Testing: Once individual modules were validated, they were combined and tested together to ensure seamless interaction and data flow.
- System Testing: The complete pipeline, from input image/video processing to final animated output, was assessed for overall functionality and reliability.
- User Feedback Evaluation: Informal user trials were conducted to gather insights on usability, visual quality, and overall satisfaction with the generated animations.

IX. CONCLUSION

This research demonstrates the effectiveness of using PCA-based motion estimation for animating still images. By moving away from traditional keypoint-based approaches, our model achieves higher realism and preserves the identity of the subject more effectively. The introduction of disentangled representation and background modeling further enhances output quality.

The system is robust, scalable, and generalizable across datasets, making it suitable for applications in education, virtual media, avatars, and telepresence. The unsupervised training strategy ensures adaptability and reduces dependency on expensive labeled data.

ACKNOWLEDGEMENT

We express our gratitude to the original authors of "Motion Representations for Articulated Animation" and to the open-source contributors whose frameworks and datasets made this research possible.

REFERENCES

[1] A. Siarohin, et al. "Motion Representations for Articulated Animation," 2021.
[2] A. Siarohin, et al. "First Order Motion Model for Image Animation," ICCV, 2019.
[3] A. Pumarola, et al. "GANimation: Anatomically-aware Facial Animation," CVPR, 2019.
[4] E. Zakharov, et al. "Few-Shot Adversarial Learning of Talking Head Models," CVPR, 2019.
[5] I. Goodfellow, et al. "Generative Adversarial Networks," NeurIPS, 2014.
[6] T. Karras, et al. "StyleGAN: A Style-Based Generator Architecture for GANs," CVPR, 2019.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi crossref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details