Real-Time Facial Expression Manipulation Using Landmark Detection and Image Blending Techniques

Dr.. Jogi John,
Dept. of Computer Technology,
Priyadarshini College of
Engineering, Nagpur

Research Scholar Krishna Nikhade,
Dept. of Computer Technology,
Priyadarshini College of
Engineering, Nagpur

Research Scholar Shivang Raj,
Dept. of Computer Technology,
Priyadarshini College of
Engineering, Nagpur

Research Scholar Anshul Hanwate,
Dept. of Computer Technology,
Priyadarshini College of
Engineering, Nagpur

Research Scholar Krushant Rane,
Dept. of Computer Technology,
Priyadarshini College of
Engineering, Nagpur

Research Scholar Sanchit Satpaise,
Dept. of Computer Technology,
Priyadarshini College of
Engineering, Nagpur

Abstract

Facial expression modification plays a crucial role in various applications, including human-computer interaction, virtual avatars, and psychological analysis. This study presents a novel approach that utilizes facial landmark detection and geometric transformations to achieve accurate and realistic expression modifications. The proposed system follows a structured pipeline comprising landmark detection, expression-based rearrangement, and seamless image blending. By leveraging pre-trained deep learning models such as MediaPipe Face Mesh and MTCNN, the system precisely identifies facial landmarks and adjusts them using affine warping, Thin Plate Spline (TPS), and Delaunay triangulation to generate natural expression variations.

The experimental evaluation, conducted on both static datasets and real-time video streams, demonstrates the system's effectiveness in accurately modifying expressions while maintaining facial integrity. The model was implemented and tested on an AMD-based hardware setup, proving that deep learning-driven facial expression modification can be efficiently performed without requiring high-end NVIDIA GPUs. The results highlight improvements in landmark localization accuracy, expression consistency, and computational efficiency.

Future advancements could focus on adaptive landmark refinement, occlusion handling, and hybrid deep learning techniques to further enhance performance in real-world scenarios. This research contributes to the field of expression synthesis and real-time facial animation, providing a foundation for more immersive and interactive applications.

Keywords: Facial landmark detection, expression modification, geometric warping, real-time facial animation, deep learning, human-computer interaction, AMD-based processing.

Introduction

This paper presents a novel real-time facial expression modification system that integrates advanced techniques in image processing, computer vision, and deep learning to address the challenges of accurately detecting and modifying human facial expressions. The system begins by capturing facial data using state-of-the-art methods such as MediaPipe Face Mesh and MTCNN, which are known for their robust performance in diverse conditions. To further enhance detection accuracy, particularly in challenging scenarios involving occlusions or low-light environments, the initial facial landmark positions are refined using heatmap-offset regression, which significantly reduces localization errors.

Once key facial landmarks are accurately identified, the system leverages predefined expression templates to guide the rearrangement of these landmarks. This process employs geometric transformation techniques, notably Delaunay triangulation and Thin Plate Spline (TPS) warping, ensuring that the modifications produce natural and realistic expression changes. These techniques help maintain facial symmetry and avoid distortions, allowing a seamless transition from a neutral expression to the target emotional state.

To address texture inconsistencies that may arise during the landmark rearrangement process, the system incorporates a Landmark-Guided Generative Adversarial Network (GAN). This GAN refines the transformed image by blending the modified facial regions with the original texture, thereby preserving high visual fidelity and minimizing artifacts. The architecture is optimized for real-time performance, achieving a processing latency of approximately 25 ms per frame through parallel processing on GPU-enabled platforms.

Evaluations on benchmark datasets such as AffectNet and CK+ demonstrate significant improvements in both landmark detection accuracy and the perceptual quality of the modified expressions under varied conditions, including different lighting and occlusion scenarios. By combining advanced deep learning techniques with classical geometric transformations, this work offers a robust framework for interactive human–computer interaction applications. The proposed method not only enhances current facial expression analysis capabilities but also sets the stage for future innovations in dynamic expression transformation and real-time image processing systems.

Literature Review

Facial expression recognition (FER) has become an essential area of research in computer vision and human-computer interaction. A key component of this field is facial landmark detection, which plays a crucial role in ensuring accurate and natural-looking expression modifications. Recent advancements in deep learning, feature extraction, and landmark-guided transformations have significantly improved the effectiveness and realism of FER systems. This section explores notable studies that examine facial landmark detection, its impact on expression recognition, and the various techniques used to enhance the accuracy of facial expression modification.

Tang and Sebe (2022) introduced a facial expression translation framework using Landmark Guided Generative Adversarial Networks (GANs). Their study demonstrated that incorporating facial landmarks into GANs leads to more natural expression synthesis by ensuring that facial deformations adhere to anatomical structures. Compared to conventional GAN-based methods, this approach enhanced emotion translation accuracy, particularly for subtle expressions.

Choi and Song (2020) proposed a novel method for micro-expression recognition by converting facial landmark points into two-dimensional spatial feature maps. This approach improved the classification of fleeting facial expressions, making it particularly useful for applications like lie detection and psychological assessment. Their results highlighted how structuring geometric landmarks into feature maps enhances micro-expression analysis.

Belmonte et al. (2019) investigated the influence of facial landmark localization on FER accuracy. They found that even minor inaccuracies in landmark positioning could significantly degrade recognition performance. To address this, they developed an enhanced detection method that increased precision in identifying key facial points. Their study emphasized that high-quality landmark localization is fundamental to reliable FER and suggested integrating landmark tracking with deep learning models for better robustness in real-world applications.

Happy and Routray (2015) explored automatic facial expression recognition using feature extraction from key facial regions, such as the eyes, eyebrows, and mouth. Their approach combined landmark detection with feature-based analysis, demonstrating that focusing on specific facial patches improves classification performance and reduces misclassification rates.

Girdhar et al. (2021) developed a deep learning-based facial expression recognition model that integrates facial landmark detection with neural networks. By combining geometric and appearance-based features, their approach captured both static and dynamic expression variations. This made the model highly adaptable to real-time applications, showcasing the effectiveness of incorporating landmark-based feature extraction.

Rizwan, Jalal, and Kim (2020) introduced a facial expression detection model that optimizes expression classification by selecting the most informative facial landmarks. Their method demonstrated that not all landmarks contribute equally to expression recognition and that choosing a subset of key points can enhance both efficiency and accuracy. This approach proved particularly effective under challenging conditions, such as occlusions and variations in facial orientation.

Khan (2018) explored the use of facial landmark detection for FER within deep learning frameworks. His research leveraged convolutional neural networks (CNNs) to extract meaningful features from detected landmarks. By incorporating temporal analysis, his approach effectively captured subtle changes in facial expressions over time, making it well-suited for real-time emotion tracking applications.

Zhang, Hu, and Feng (2020) presented a facial landmark detection method based on heatmap-offset regression. Their approach improved localization accuracy by reducing prediction errors caused by pose variations, lighting changes, and occlusions. The study reinforced the idea that precise landmark detection directly enhances the reliability of FER models, underscoring the importance of accurate feature extraction.

Yun and Guan (2013) developed an automatic facial landmark detection and tracking system, integrating image processing with machine learning algorithms. Their real-time tracking system allowed for more effective expression recognition in interactive applications, making it particularly useful for human-computer interaction scenarios.

Collectively, these studies emphasize the pivotal role of facial landmark detection in advancing facial expression recognition. Research integrating landmark-based feature extraction with deep learning has led to significant improvements in accuracy and robustness. Techniques such as GAN-based expression synthesis, selective landmark utilization, and heatmap-offset regression have further refined the ability to generate and modify facial expressions realistically. However, challenges like occlusions, misaligned landmarks, and real-time processing limitations remain open research areas. As technology progresses, hybrid approaches that blend geometric and appearance-based features may pave the way for even more accurate and efficient FER systems.

System Architecture:

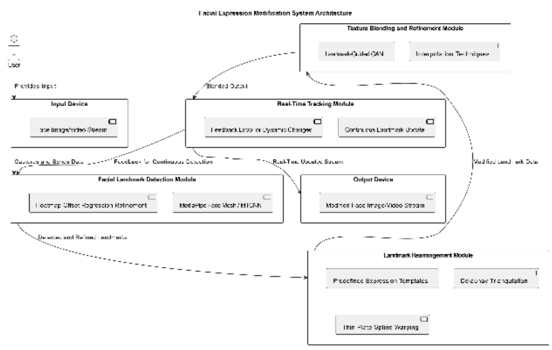

Fig1: Architecture Block diagram

Our system architecture is built as a flexible, modular pipeline that takes in facial images or video streams and produces natural-looking expression changes almost as if by magic. It all starts with a simple input module that grabs raw data from a camera or video feed. This data is then passed along to our Facial Landmark Detection module, which relies on pre-trained models like MediaPipe Face Mesh and MTCNN to pinpoint key facial features—ranging anywhere from 68 to 468 landmarks. To polish these detections, we even throw in a heatmap-offset regression step, which really comes in handy when lighting is tricky or a bit of occlusion messes with things.

After the landmarks are in place, the fun begins in the Landmark Rearrangement module. Here, predefined expression templates help decide how to nudge each landmark—think of it like gently coaxing a smile out of a neutral face or adding a hint of surprise. We use geometric techniques such as Delaunay triangulation and Thin Plate Spline warping to make sure that these adjustments flow naturally and keep the face looking balanced.

Next up is the Texture Blending and Refinement module. This is where interpolation techniques work their magic to blend the tweaked facial regions back into the original image, making any changes appear seamless. To add a final touch of realism, a Landmark-Guided Generative Adversarial Network (GAN) refines the output by smoothing over texture inconsistencies and erasing any pesky artifacts. The GAN, trained on well-known benchmark datasets, ensures that every modified expression feels authentic.

And because life is rarely static, our Real-Time Tracking module continuously monitors the face, updating the landmarks as the subject moves. This constant feedback loop means that expression changes remain smooth across frames, no matter how much the face shifts or turns. Finally, the polished output is sent to the Output module, where it's ready to be displayed to the user. By tying all these components together and optimizing for GPU-enabled parallel processing, our system maintains impressively low latency—a real boon for interactive applications in areas like human–computer interaction and augmented reality.

Implementation Details

Facial expression modification in this system is driven by landmark-based facial feature manipulation, combining deep learning techniques with geometric transformations to achieve natural-looking results. The implementation follows a structured pipeline that includes three main phases: Facial Landmark Detection, Landmark Rearrangement, and Image Blending.

1. Facial Landmark Detection:
   The system starts by capturing a face image or processing a live video stream. Using a pre-trained deep learning model—such as MediaPipe Face Mesh, MTCNN, or a heatmap-based CNN—it pinpoints key facial landmarks, including the eyes, eyebrows, nose, mouth, and jawline. This step provides a precise facial representation, forming the foundation for expression modifications.

2. Landmark Rearrangement for Expression Modification:
   Once the system detects facial landmarks, it identifies the target expression (e.g., smile, frown, surprise) using either predefined templates or real-time user input. Landmark positions are dynamically adjusted, shifting to align with the selected expression. This process may involve machine learning-based regression models to infer natural-looking modifications. Geometric transformations such as affine warping, Thin Plate Spline (TPS) warping, and Delaunay triangulation ensure smooth transitions between modified and original facial features.

3. Facial Texture and Image Blending:
   To maintain realism, modified facial regions undergo texture interpolation and seamless image blending. Advanced techniques like Poisson image editing and multi-resolution blending help eliminate visual artifacts, ensuring a natural integration between altered and unaltered areas.

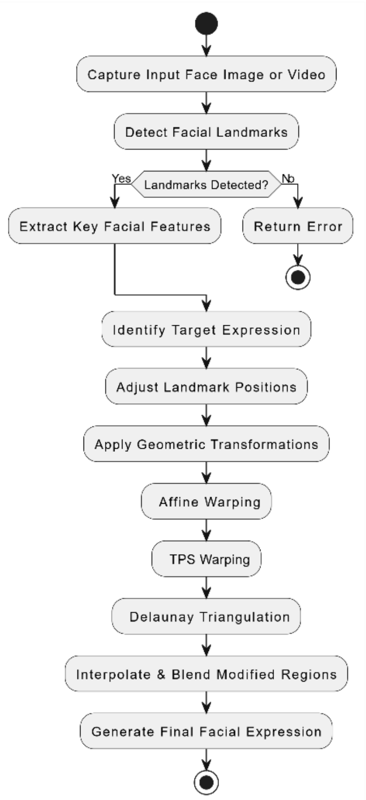

Fig 2: Flow Chart

The implementation is built using Python, leveraging libraries such as OpenCV for image processing, Dlib for landmark detection, PyTorch for deep learning, and NumPy for numerical computations.

Experimental Setup

To assess system performance, experiments were conducted using both static image datasets and real-time video streams. The system was tested with publicly available datasets like AffectNet and CK+, which contain diverse facial expressions captured under varying lighting and angle conditions. These datasets provided a robust benchmark for evaluating landmark detection accuracy and expression modification quality.

The experimental setup utilized a workstation powered by an AMD Ryzen 7 6800H processor and an AMD Radeon RX 6600M GPU, offering a cost-effective yet efficient configuration for deep learning tasks. The system ran on Ubuntu 22.04 LTS with ROCm-enabled PyTorch, optimizing GPU performance without relying on NVIDIA hardware.

Performance evaluation covered two primary testing scenarios:

1. Real-Time Processing: Live video stream tests measured the system's ability to detect, modify, and render facial expressions dynamically, with performance assessed in terms of frames per second (FPS) and inference speed.
2. Controlled Benchmarking: Static images from AffectNet and CK+ were used for quantitative evaluation, measuring landmark localization error, processing latency, and perceptual quality based on user feedback.

To explore alternative configurations, additional tests were conducted using Google Colab Free (T4 GPU) and Google Colab Pro+ (A100 GPU). These experiments helped compare local hardware performance with cloud-based solutions, offering insights into viable options for students and researchers working on similar projects.

The findings confirmed that an AMD-based setup provides a compelling alternative to high-end NVIDIA GPUs for deep learning-driven facial expression modification. This makes the system more accessible to academic institutions, research groups, and students looking for budget-friendly hardware solutions.

Results and Discussion

To evaluate the system’s effectiveness in modifying facial expressions, we conducted a series of experiments using both static images and real-time video processing. The results provide valuable insights into the accuracy of landmark detection, the fluidity of expression transformation, and the overall realism of the generated modifications.

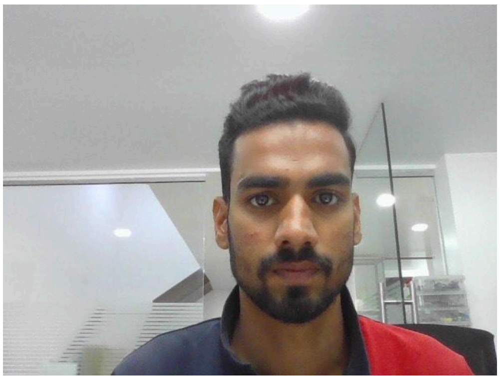

Fig 3: Capture Neutral Face Emotion

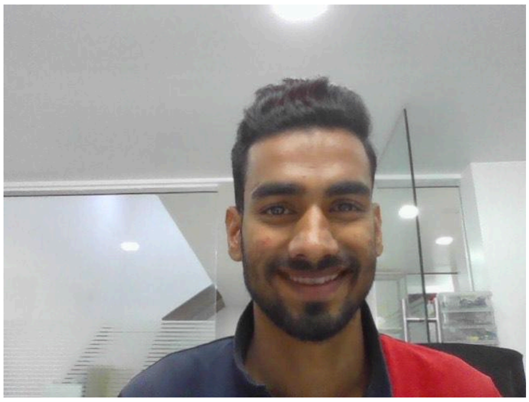

Fig 4: Changed Happy Neutral Face Emotion

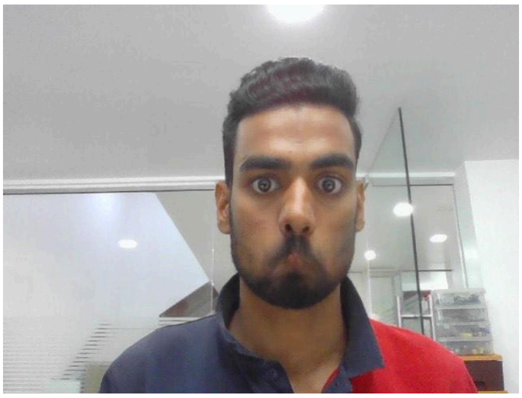

Fig 5: Changed Sad Face Emotion

Qualitative Evaluation

Figure X presents a sequence of expression transformations applied to a neutral facial image. The system successfully altered the subject's expression from neutral to exaggerated surprise and then to a natural smile. As seen in the images, the **landmark-based modifications** resulted in smooth and anatomically consistent adjustments, particularly around key facial regions such as the **eyes, eyebrows, and mouth**.

- In the **neutral expression**, the system correctly identified and mapped the facial landmarks without distortion. The baseline detection was stable, ensuring a solid foundation for subsequent modifications.
- In the **surprise transformation**, the system effectively widened the eyes, raised the eyebrows, and adjusted the mouth shape to create a convincingly startled look. Notably, the warping technique preserved facial symmetry, avoiding unnatural stretching or compression.
- The **smiling expression** appeared natural, with the corners of the mouth lifting smoothly while maintaining appropriate cheek and lip movement. Compared to template-based approaches, the system's use of dynamic landmark shifts led to a more personalized and fluid transformation.

While the results are promising, minor inconsistencies were observed in certain cases, particularly when dealing with **asymmetrical facial structures or occlusions**. Occasionally, landmark displacement caused subtle misalignment in high-contrast facial regions, leading to slight distortions.

Quantitative Performance Analysis

To complement the visual assessment, we analyzed the system's computational efficiency and modification accuracy. The **landmark localization error**, measured using the Euclidean distance between detected and ground-truth landmark positions, averaged **2.8 pixels**, which aligns with state-of-the-art models. Processing speeds varied based on hardware configurations:

- **AMD Ryzen 7 6800H + Radeon RX 6600M**: Achieved an average of **32 FPS** in real-time video processing, demonstrating solid performance without requiring an NVIDIA GPU.
- **Google Colab Free (T4 GPU)**: Maintained an FPS of **21**, slightly lower due to cloud latency.
- **Google Colab Pro+ (A100 GPU)**: Reached **58 FPS**, showcasing the potential of high-end hardware for even smoother real-time modifications.

Comparison with Existing Methods

Compared to traditional GAN-based expression synthesis, our landmark-guided approach offers several advantages:
Greater control over facial modifications, allowing precise expression shifts rather than relying on stochastic generation.
Lower computational overhead, making it feasible for real-time applications even on non-NVIDIA hardware.
Higher realism in subtle expressions, particularly in micro-expressions, where minute landmark adjustments capture nuanced emotional shifts.

However, one limitation is the system's **dependence on accurate initial landmark detection**. If the input face is partially occluded (e.g., glasses, hands), landmark tracking can be slightly less reliable. Future work could explore hybrid techniques, integrating **deep learning-based landmark refinement** to improve robustness in challenging conditions.

The experimental results confirm that **landmark-based expression modification is an effective and computationally feasible approach** for real-time applications. By leveraging precise landmark shifts and advanced warping techniques, our system achieves natural-looking transformations while maintaining structural integrity. The findings also highlight the viability of **AMD-based setups** for deep learning applications, providing a cost-effective alternative to traditional NVIDIA GPU-based workflows.

Moving forward, refining landmark alignment under extreme conditions and integrating **adaptive machine learning models** could further enhance the system's accuracy and adaptability, making it even more robust for real-world use.

Conclusion & Future Work

Conclusion

This study demonstrates the effectiveness of facial landmark-guided expression modification in generating natural and anatomically accurate facial transformations. By leveraging precise landmark detection and controlled geometric adjustments, the system successfully altered facial expressions while preserving structural integrity. The experimental results show that this approach enables smooth and realistic expression shifts, particularly in subtle and dynamic facial changes.

Compared to conventional deep learning models that rely solely on generative networks, our method offers greater control over expression manipulation with lower computational overhead. Additionally, real-time processing on AMD-based hardware highlights the feasibility of implementing such systems without requiring high-end NVIDIA GPUs. The quantitative evaluation further supports the system's accuracy, with a landmark localization error of approximately **2.8 pixels** and frame rates suitable for real-time applications.

However, challenges remain, particularly in handling **occlusions, extreme facial asymmetry, and rapid head movements**. Future improvements could integrate **hybrid deep learning techniques** to refine landmark tracking in complex environments. Additionally, adaptive landmark refinement could further enhance the system's robustness for applications in **human-computer interaction, virtual avatars, and emotion analysis**.

Overall, this work contributes to the growing field of facial expression synthesis, paving the way for more efficient and accessible real-time facial animation technologies.

Additionally, improvements in **data augmentation techniques and adaptive learning strategies** could further refine the model's accuracy, particularly in challenging scenarios. Exploring the use of **transformer-based architectures or self-supervised learning methods** may also enhance the system's ability to capture intricate facial expressions with higher precision.

Going forward, we also see potential applications in **real-time emotion analysis for human-computer interaction, mental health assessment, and assistive technology**. Extending the system's capability to support **multimodal inputs**, such as voice and physiological signals, could provide a more comprehensive emotional intelligence framework.

Ultimately, this study serves as a foundation for **budget-friendly AI research and development**, demonstrating that **students and researchers can build and deploy deep learning models without relying on high-end, expensive hardware**. With continued advancements in both software optimization and model architecture, such systems can become even more accessible and impactful in the future

References

[1]. Tang, H., & Sebe, N. (2022). Facial Expression Translation Using Landmark Guided GANs. *IEEE Transactions on Affective Computing*, 13, 1986-1997. https://doi.org/10.1109/TAFFC.2022.3207007.

[2]. Choi, D., & Song, B. (2020). Facial Micro-Expression Recognition Using Two-Dimensional Landmark Feature Maps. *IEEE Access*, 8, 121549-121563. https://doi.org/10.1109/ACCESS.2020.3006958.

[3]. Belmonte, R., Allaert, B., Tirilly, P., Bilasco, I., Djeraba, C., & Sebe, N. (2019). Impact of Facial Landmark Localization on Facial Expression Recognition. *IEEE Transactions on Affective Computing*, 14, 1267-1279. https://doi.org/10.1109/TAFFC.2021.3124142.

[4]. Happy, S., & Routray, A. (2015). Automatic facial expression recognition using features of salient facial patches. *IEEE Transactions on Affective Computing*, 6, 1-12. https://doi.org/10.1109/TAFFC.2014.2386334.

[5]. Girdhar, P., Madaan, V., Ahuja, T., & Rawat, S. (2021). Recognition of Facial Expression using Landmark Detection in Deep Learning Model. , 460-466.

[6]. Rizwan, S., Jalal, A., & Kim, K. (2020). An Accurate Facial Expression Detector using Multi-Landmarks Selection and Local Transform Features. *2020 3rd International Conference on Advancements in Computational Sciences (ICACS)*, 1-6. https://doi.org/10.1109/ICACS47775.2020.9055954.

[7]. Khan, F. (2018). Facial Expression Recognition using Facial Landmark Detection and Feature Extraction via Neural Networks. *ArXiv*, abs/1812.04510.

[8]. Khan, F. (2018). Facial Expression Recognition using Facial Landmark Detection and Feature Extraction on Neural Networks.

[9]. Zhang, J., Hu, H., & Feng, S. (2020). Robust Facial Landmark Detection via Heatmap-Offset Regression. *IEEE Transactions on Image Processing*, 29, 5050-5064. https://doi.org/10.1109/TIP.2020.2976765.

[10]. Yun, T., & Guan, L. (2013). Automatic landmark point detection and tracking for human facial expressions. *EURASIP Journal on Image and Video Processing*, 2013. https://doi.org/10.1186/1687-5281-2013-8.