DOI: https://doi.org/10.36094/sc.v92.2026.Development_of_an_AI_Based_Image_Slideshow.Khatun.290

ARTICLE

Sci. and Cult. 92 (3–4) : 290-292 (2026)

DEVELOPMENT OF AN AI-BASED IMAGE SLIDESHOW VIDEO CREATOR USING DEEP LEARNING

BEAUTY KHATUN[1]*, DEBAYAN PANDIT[1], SANJIB GHOSH[1], SUVANKAR BARAI[1] AND RAJIB CHAKRABARTY[2]

The proliferation of digital photography and social media platforms has amplified the demand for automated slideshow video generation from extensive image repositories. Conventional slideshow software, however, relies on templates and necessitates considerable manual intervention. This study introduces an AI-powered system for creating image slideshow videos, leveraging deep learning methodologies. The system autonomously filters out low-quality and duplicate images, employs Convolutional Neural Networks to extract visual features, generates image captions via vision-language models, and predicts the overall mood to recommend suitable music and transitions. Empirical findings derived from real-world photo albums demonstrate enhanced video coherence, diminished user workload, and greater user satisfaction when contrasted with traditional slideshow creation approaches.

Existing systems are not content-aware, depend heavily on static templates, and fail to generate emotion-aligned story- telling videos automatically. Therefore, an intelligent solution is required that generates slideshow videos automatically using deep learning.

The key objectives of this research include:

1) Design an automated pipeline for slideshow video gen- eration from image collections.
2) Use deep learning for feature extraction, image under- standing, and caption generation.
3) Implement image quality assessment and duplicate re- moval to improve slideshow clarity.
4) Develop an intelligent ordering mechanism for story- telling.
5) Recommend transitions and music based on emotion mood detection.
6) Render slideshow videos with optimized processing and minimal user effort.

Related Work

Deep learning models such as ResNet, EfficientNet, and MobileNet have been widely used for image classification and feature extraction due to their strong visual representa- tion capabilities. Captioning techniques evolved from CNN- RNN methods to transformer-based vision-language models, improving caption accuracy and fluency. Emotion recognition in images often uses facial expres- sion recognition based on CNN classifiers. Some works also fully integrated frameworks that combine selection, ordering, captioning, and mood-aware rendering into an end-to-end slideshow creation pipeline remain limited.

Methodology

Proposed System: **System Overview:** The proposed AI slideshow creator consists of the following modules:

1) Image input and preprocessing
2) Deep feature extraction using CNN
3) Image quality assessment
4) Duplicate image detection and removal
5) Clustering and storyline ordering
6) Caption generation
7) Mood-based music and transition recommendation
8) Video rendering and export.

Architecture: The architecture is shown in Fig. 1.

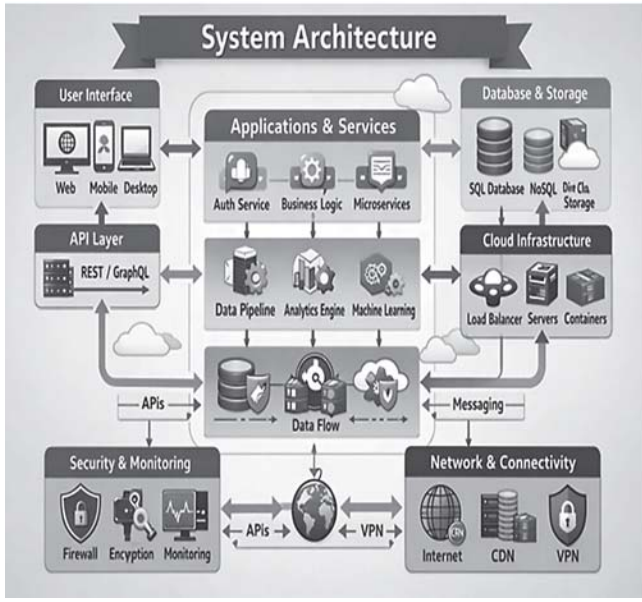

Fig.1 System architecture of the proposed AI-based image slideshow video creator.

Image Preprocessing: Each image is resized and normalized to a standard resolution. Contrast enhancement and noise reduction improve robustness. EXIF metadata is extracted when available.

Feature Extraction: A pretrained CNN extracts embeddings from images:

$$Fi = CNN\ (Ii) \quad (1)$$

where Ii is the input image and Fi is the corresponding feature vector.

Image Quality Assessment: Blur detection uses Laplacian variance:

$$B = V\ ar(\nabla\ 2I) \quad (2)$$

Images with blur scores below a threshold are removed.

Duplicate Detection: Cosine similarity measures redundancy:

$$Sim(Fa, Fb) = \frac{Fa \cdot Fb}{||F\ ||\ ||F\ ||}$$

If Sim > 0.95, one image is considered duplicate and discarded.

Storyline Ordering: Ordering strategy:

- primary: timestamp sorting
- secondary: semantic clustering
- within-cluster: nearest-neighbor traversal

This improves narrative flow and coherence.

Caption Generation: A vision-language model

generates captions for each selected image. Captions can be inserted as subtitles or overlay text.

Emotion Detection and Recommendation: Emotion categories:

- Happy
- Sad
- Calm
- Excited

Music is selected from a mood-tagged library and transitions are applied based on scene variation and emotion.

Video Rendering: Selected images are converted into a timed sequence with transitions and synchronized audio, rendered using FFmpeg or MoviePy.

Algorithm: Algorithm 1: AI Slideshow Video Generation Input: *Album S = {I1, I2, . . . , In}*

Output: *Slideshow video V*

1) Load image set S
2) Preprocess images (resize, normalize)
3) Extract features Fi using CNN
4) Compute quality score and remove low-quality images
5) Compute cosine similarity and remove duplicates
6) Cluster images using embeddings
7) Order images using timeline + cluster ordering
8) Generate captions for ordered images
9) Detect mood and recommend music + transitions
10) Render slideshow video using video engine
11) Return output video V

Implementation Details: **Software Tools:** The system is implemented using:

- Python 3.x
- OpenCV (preprocessing)
- TensorFlow/PyTorch (deep models)
- FFmpeg/MoviePy (video rendering)

Hardware Requirements: Minimum requirements:

- CPU: Intel i5 or above
- RAM: 8 GB
- GPU: optional (reduces processing time)

Results and Discussion

Experiments were performed on real-world albums such as travel, birthday, and college programs. The system produced coherent and emotion-aligned slideshow videos with reduced repetitive images. TABLE. 1 presents a comparison between traditional tools and the proposed system.

Table 1: Comparison Of Traditional Slideshow Tools Vs Proposed System

Feature	Traditional Tool	Proposed System
Image selection	Manual	Automatic
Duplicate handling	No	Yes
Captioning	Limited	Automatic
Emotion music	No	Yes
Story coherence	Medium	High
User effort	High	Low

Applications

The proposed system can be applied in:

- social media content creation
- event photography (weddings, birthdays)
- academic presentations and college programs
- travel memory compilation
- digital marketing storytelling

Conclusion And Future Work

This paper presented the design and development of an AI-based image slideshow video creator using deep learning. The system automatically selects high-quality images, removes duplicates, orders images meaningfully, generates captions, recommends music and transitions based on emotion, and renders a complete slideshow video. It significantly reduces manual effort while improving storytelling quality compared to template-based tools.

Future extensions may include:

1) narration generation using Text-to-Speech
2) automatic beat synchronization with music
3) multilingual subtitles
4) mobile deployment using lightweight models
5) generative AI-based music creation ❒

References

1. K. He, X. Zhang, S. Ren, and J. Sun, *in Proc. CVPR,* 2016.
2. A. Vaswani et al., in *NeurIPS,* 2017.
3. O. Vinyals et al., in *Proc. CVPR,* 2015.
4. OpenCV Documentation, "Open Source Computer Vision Library," 2024. [Online]. Available: https://opencv.org/
5. FFmpeg Documentation, "FFmpeg Multimedia Framework," 2024. [On- line]. Available: https://ffmpeg.org/
6. S. Li and W. Deng, "Deep Facial Expression Recognition: A Survey," IEEE Transactions on Affective Computing, 2022.