International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 01 | Jan - 2025 SJIF Rating: 8.448 ISSN: 2582-3930

Human Identification System Using Videos

Sanjana N[1], Sanjay S[2], Surya R[3], Usman Ayan Ulla Khan Suri H J[4]

[1]Sanjana N, Department of ISE, JNN College of Engineering.
[2]Sanjay S, Department of ISE, JNN College of Engineering.
[3]Surya R, Department of ISE, JNN College of Engineering.
[4]Usman Ayan Ulla Khan Suri H J, Department of ISE, JNN College of Engineering.

---***---

Abstract - The present study focuses on the challenges of human long-range recognition with deep learning, and AlexNet performance is considered, being one of the most recognized convolutional neural networks. A specially curated dataset was used to mimic the real world in a more practical and realistic scenario, which involves subjects varying in distances and angles while creating low-resolution issues with variability in lighting and distractions from the environment. The system starts by passing the input dataset through a local Laplacian filter that is utilized to enhance the quality of images and videos before feeding it into a pre-trained AlexNet model. The results reveal that AlexNet reaches a recognition accuracy of 79-80% and shows that image properties such as sharpness and contrast significantly affect performance, as measured by the mean and standard deviation. The study underlines requirements of high-quality images for applications in long-range human recognition and sheds insight into the ways enhancements in images and videos could improve performance in deep learning models. Results support further evolution of more potent human recognition systems, particularly in application domains such as surveillance and security in which the accurate identification of a target from a long distance is highly critical.

***Key Words*:** Long-range recognition, Human recognition, Deep learning, Image and Video processing.

1. INTRODUCTION

Long-range human recognition, the task of identifying individuals from images and videos captured at significant distances, has become an increasingly critical area of study and application, especially in fields such as surveillance, autonomous vehicles, and security systems. As technology advances, the collection of visual data has grown exponentially, highlighting the need for robust and efficient methods to process and analyze images that capture people from far away. Human recognition in such scenarios is a crucial component for various security and monitoring systems where accurate identification from a distance can have significant implications for public safety, law enforcement awareness.

One of the most significant challenges in long-range human recognition lies in the limitations imposed by the quality of the images captured from a distance. Unlike close-range images where individuals are more easily distinguishable, long-range images are affected by several factors such as resolution degradation, motion blur, occlusion, and varying lighting conditions. These factors reduce the size and clarity of human figures in the image and video, making it much harder to accurately identify individuals. Additionally, the reduced size of subjects, often combined with complex background environments, further complicates feature extraction and analysis, presenting a unique set of challenges compared to traditional close-range human recognition tasks.

The growing demand for effective long-range human recognition methods has led to significant advancements in deep learning, particularly with the development of convolutional neural networks (CNNs). CNNs have become the backbone of modern computer vision techniques, enabling automated systems to perform tasks that were previously unimaginable, such as object detection, image segmentation, and face recognition. One of the most influential models in this field is AlexNet, a pioneering deep learning architecture that has demonstrated remarkable success in visual recognition tasks. AlexNet's ability to extract features from images, even under challenging conditions, has made it a valuable tool in various applications. However, its effectiveness in the context of long-range human recognition remains an area that has not been extensively explored, and it presents an opportunity for further investigation.

While AlexNet excels in controlled environments with high-resolution images, long-range images and videos often suffer from a variety of quality issues, such as low resolution, noise, and poor contrast, making it difficult for the network to distinguish important features. This research aims to assess AlexNet's performance in recognizing humans from images captured at varying distances and under diverse environmental conditions. By focusing on key image quality factors such as sharpness, brightness, contrast, and noise, the study seeks to understand how these attributes impact the accuracy of human recognition models in challenging long-range scenarios. In doing so, the paper examines the potential of AlexNet for real-world applications where traditional human recognition systems, which rely on high-resolution, close-range images, fall short.

The study also aims to inform future developments in human recognition technologies, contributing to the design of systems capable of reliably identifying individuals from a distance. As security, surveillance, and autonomous navigation systems become more integrated into our daily lives, the ability to recognize human subjects in dynamic, long-range environments will be of paramount importance. By analyzing the performance of AlexNet on a custom dataset that includes both high- and low-quality images, this research provides valuable insights into the strengths and limitations of deep learning models in these types of recognition tasks. Ultimately, the findings have the potential to guide the creation of more accurate, efficient, and adaptable recognition systems that can meet the demands of modern security and monitoring applications.

2. LITERATURE REVIEW

A review by Patrycja Miazek et al. (2024) on the usage of radar data in the context of "Human Activity Recognition", demonstrated that such a solution is non-invasive and specifically privacy-preserving in scenarios with low lighting or penetration through walls. A review is given of such techniques as micro-Doppler analysis, deep mopdels, for example, CNNs, RNNs, LSTMs and systems, for example, RadHAR with 90.47% accuracy and Mobile-RadarNet being a lightweight edge system aiming at efficiency in computational resources. Although the paper focuses on the advantages of using LiDAR and data fusion with radar, multi-person recognition challenges and scalability issues will prevail; hence, more research should be conducted to enhance the HAR system.

Christian Gianoglio et al. (2024) proposed a human action recognition system for radar using deep learning on a Raspberry Pi 4. The idea combines FMCW Range-Doppler maps with a DNN that uses CNNs in spatial features and LSTMs for temporal dependencies, offering an accuracy of 93.2% in multiclass and 96.8% in binary classification regarding actions such as walking, sitting, and falling. It works efficiently with a 4% false-negative rate for harmful actions and displays real-time performance with ~2.95 second inference times. The system is ideal to deal with indoor safety and elderly fall detection, ensures privacy and works well in cluttered environments. Future improvements could extend its use to multiple targets and more dynamic scenarios.

A. A. Alzahrani's paper of 2024 has presented BIPFER-EOHDL, a facial expression recognition system using deep learning techniques such as EfficientNetB7 for feature extraction and MA- BLSTM for classification. The accuracy is 99.05%. This model outperforms models like ResNet and VGGNet. The results are promising, but the real-time performance on larger datasets is untested. The system is recommended for application in healthcare, emotional well-being monitoring, and human- computer interaction.

According to Su Keun Jeong et al. (2023), it is possible to implement human activity recognition using a hybrid scheme with joint angle changes, accompanied by a 3D-deep convolutional neural network (DCNN). This model differs from previous approaches because it might have an opportunity to differentiate between even comparable movements like standing or sitting by focusing on changing the angle. The model is very successful in recognising the human activities with an accuracy over 96% on the datasets similar to UTKinect Action3D. Though it has several problems computationally when applied to vast video datasets, this approach indeed shows promise in real applications in smart homes and elder monitoring systems.

Jag Mohan Singh et al. (2024) researches the spoofing attack of composite face images on Face Recognition Systems (FRS) created by GANs. The authors produced a dataset of 500,000 images and proposed the Generalized Morphing Attack Potential (G-MAP) metric for quantifying the vulnerability of FRS. It illustrates how, under the generation of GAN, the face can counter the traditional FRS. They need more sophisticated deep models for enhancing the robustness of FRS to adaptive spoofing techniques.

Taeseong Gim and Kyung-Ah Sohn (2024) Introduces a technique with incorporation of Noise Direction Regularization to enhance face recognition over the noisy and low-resolution setup. This enhances training of the model beyond regular loss functions like CosFace and AdaFace through noise sample identification as well as feature norms. Still promising for tasks such as TinyFace and SurvFace, it requires fine-tuning for better efficiency and practical handling of noisier data for achieving better scalability.

Sathisha Basavaraju (2024) proposes a human action recognition system for aerial surveillance that is using Elliptical Modeling with deep learning. It reduces the distortion created in camera angles and altitudes during image capture, leading to higher accuracy in detection. Through integration of CNNs and RNNs, spatial and temporal features are incorporated and therefore improve performance for scenarios of crowded environments. Yet the challenges remain in addressing issues like occlusion and scalability;

therefore, such methods still require further developments in the real-world applications towards public safety.

Bum-soo Kim and Sanghyun Seo (2022) developed an intelligent agent system applying deep learning- based face recognition and digital human interfaces for personalized services, which is adjusted based on the user's preference. Data processing and keyframe interpolation are applied to enhance user interaction through this system. It has been verified through facial recognition with CutMix to enhance accuracy as well as service delivery within the academe, and user reception was satisfactory whereby 81% of users reported having found it helpful. As such, the authors advise continued development using complex algorithms like Graph Neural Networks for higher robustness and interpersonal relation handling.

A paper presented by Madhusree Das et al. in 2024 explores the classical, hybrid and deep learning-based face recognition techniques. With such great achievements in the management of robust facial variations, deep learning-based models gained prominence in most of its research works including CNNs. Nevertheless, such issues concerning lighting, pose, ageing, and privacy continue. Future studies might require additional development to make these strong models along with enhanced performance while working in real-time together with considerations for huge-volume responsible use.

3. PROPOSED METHEDOLOGY

The proposed methodology for long-range human recognition leverages a combination of deep learning models and image enhancement techniques to address the unique challenges posed by long-range images and videos, including low resolution, varied lighting conditions, and environmental distractions.

Data Collection and Dataset Preparation:
A diverse dataset consisting of images and videos captured at different distances, angles, and environmental conditions is curated. The dataset includes human subjects captured at various ranges to simulate real- world scenarios such as surveillance, security, and crowd monitoring. Images and videos are annotated with the corresponding class labels for human identification.

Preprocessing:
To address issues such as noise and low resolution, preprocessing techniques are employed. A local Laplacian filter is used to enhance local structural features and improve image sharpness. This step ensures that important details such as human figures, even at a distance, are better represented in the input data.

Feature Extraction using Convolutional Neural Networks (CNN):
The preprocessed images are fed into a pre-trained deep learning model, such as AlexNet, a CNN architecture known for its success in image classification tasks. The CNN is fine-tuned on the curated dataset to extract relevant features for human recognition. This step focuses on extracting spatial features from the images to identify human figures in challenging conditions.

Model Training:
The model is trained using a combination of supervised learning techniques. The training process is carried out using a training dataset with annotated human figures, applying techniques like data augmentation to further improve model robustness. The network learns to classify the human subjects by their unique features, despite being captured from long-range distances.

Evaluation and Performance Metrics:
The performance of the model is evaluated using a set of key metrics such as accuracy, precision, recall, F1- score, and confusion matrix. Special attention is given to the ability of the model to recognize humans in images with various challenges like low resolution and poor lighting. The sharpness of the images and their mean and standard deviation values are also considered as part of the evaluation to understand the impact of image quality on recognition success.

Model Optimization:
Based on the evaluation results, the model undergoes further optimization, which may involve adjusting hyperparameters or implementing transfer learning techniques to improve performance. This stage may also explore alternative architectures such as ResNet or VGGNet if needed, to compare and determine which model yields the best results in long-range recognition scenarios.

Real-World Application Testing:
After training and optimization, the model is tested on real-world datasets or in real-time settings where human recognition is required in long-range situations (e.g., outdoor surveillance or crowd monitoring). This phase assesses the model's adaptability to varying environmental conditions, such as outdoor lighting, occlusions, and crowded scenes.

Final Evaluation and Recommendations:
The final model performance is analyzed, and recommendations are made for further improvements. Potential future work includes integrating the model into automated surveillance systems for human tracking and implementing edge computing techniques to reduce computational load and improve real-time processing capabilities in real-world applications.

4. SYSTEM ARCHITECTURE

The system architecture for long-range human recognition is designed to process, recognize, and track human figures in long-range surveillance scenarios, such as aerial surveillance or security camera feeds. It integrates various components to handle the complexities of detecting and recognizing humans at significant distances while dealing with issues like occlusion, low resolution, and varying environmental conditions. The architecture follows a modular structure and includes the following:

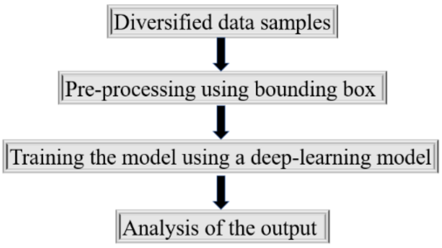

Fig 1: Human recognition framework

The components of the proposed system are explained below:

Data Collection and Data Samples:
The system begins with collecting human action images and videos through surveillance cameras, drones, or other long-range observation devices. These images are captured at various distances, resolutions, and angles, ensuring diversity in the dataset. A variety of environments, such as indoor and outdoor scenes, are included to improve the model's ability to generalize across different settings.

Data Preprocessing:
Once the images and videos are collected, they undergo a series of preprocessing steps to improve their quality and prepare them for model input. This involves resizing images to a standard size, normalizing pixel values to a consistent range, and applying data augmentation techniques such as rotation, flipping, and cropping. These steps help to improve model robustness and prevent overfitting by artificially expanding the dataset.

Bounding Box Detection:
For accurate human action recognition, the images are processed to locate and isolate the human subjects. This is done by applying a bounding box detection algorithm, such as YOLO (You Only Look Once), which is trained to identify and segment humans from the background. The bounding box isolates the region of interest, ensuring that human actions are the primary focus of the analysis.

AlexNet:
The cropped images within the bounding boxes are then fed into AlexNet, a pre-trained Convolutional Neural Network (CNN), for feature extraction and classification. AlexNet's deep architecture is leveraged to automatically detect and classify complex patterns in the images, distinguishing between different human actions such as walking, sitting, or standing.

Analysis of the Output:
After the human actions are classified, the output is evaluated based on accuracy and other performance metrics. The results are analyzed to identify any misclassifications or areas of improvement. The system is then fine-tuned to optimize its performance, ensuring high recognition accuracy even in challenging real-world scenarios like low-light conditions or occlusion.

5. RESULTS

The results of the project demonstrate significant improvements in human recognition through the application of image processing techniques. The Bound Box technique effectively isolated key features, such as the human figure, improving the focus of the recognition algorithm and reducing computational overhead. Local Laplacian Filtering enhanced the sharpness, contrast, and detail of the images and videos, making it easier for the AlexNet model to extract relevant features, leading to better classification results. The filtered dataset showed consistent improvements in sharpness and contrast, addressing challenges such as sparse features and variations across poses. A property assessment after filtering revealed significant enhancements in mean, covariance, standard deviation, and sharpness, validating the effectiveness of Laplacian filtering in refining the dataset. These findings highlight the success of the applied techniques in improving image quality and the overall performance of the human recognition system.

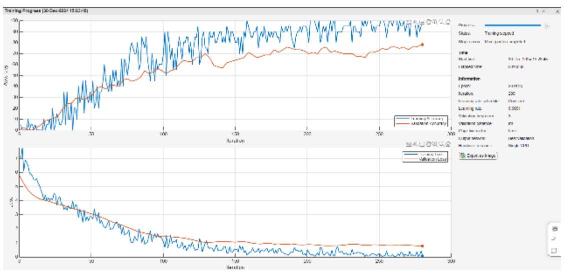

Fig 2: Accuracy of the training

6. CONCLUSION

By applying Bound Box for feature isolation and Local Laplacian filtering for image enhancement, the model's accuracy improved. The refined dataset, combined with AlexNet, provided more reliable human recognition, demonstrating the importance of these preprocessing techniques in complex recognition tasks.

REFERENCES

1. E. S. Rahayu, E. M. Yuniarno, I. K. E. Purnama and M. H. Purnomo, "A Combination Model of Shifting Joint Angle Changes With 3D-Deep Convolutional Neural Network to Recognize Human Activity," in IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 32, pp. 1078-1089, 2024, doi: 10.1109/TNSRE.2024.3371474.

2. C. Gianoglio, A. Mohanna, A. Rizik, L. Moroney and M. Valle, "On Edge Human Action Recognition Using Radar-Based Sensing and Deep Learning," in IEEE Transactions on Industrial Informatics, vol. 20, no. 3, pp. 4160-4172, March 2024, doi: 10.1109/TII.2023.3316164.

3. P. Miazek, A. Żmudzińska, P. karczmarek and A. Kiersztyn, "Human Behavior Analysis Using Radar Data: A Survey," in IEEE Access, vol. 12, pp. 153188-153202, 2024, doi: 10.1109/ACCESS.2024.3474100.

4. H. Hofbauer, F. Autrusseau and A. Uhl, "Low Quality and Recognition of Image Content," in IEEE Transactions on Multimedia, vol. 24, pp. 3595-3610, 2022, doi: 10.1109/TMM.2021.3103394.

5. A. A. Alzahrani, "Bioinspired Image Processing Enabled Facial Emotion Recognition Using Equilibrium Optimizer with a Hybrid Deep Learning Model," in IEEE Access, vol. 12, pp. 22219-22229, 2024, doi: 10.1109/ACCESS.2024.3359436.

6. J. Park, J. Kim, Y. Gil and D. Kim, "DGU-HAO: A Dataset with Daily Life Objects for Comprehensive 3D Human Action Analysis," in IEEE Access, vol. 12, pp. 8780-8790, 2024, doi: 10.1109/ACCESS.2024.3351888.

7. U. Azmat et al., "An Elliptical Modeling Supported System for Human Action Deep Recognition Over Aerial Surveillance," in IEEE Access, vol. 11, pp. 75671-75685, 2023, doi: 10.1109/ACCESS.2023.3266774.

8. J. M. Singh and R. Ramachandra, "Deep Composite Face Image Attacks: Generation, Vulnerability and Detection," in IEEE Access, vol.11, pp. 76468-76485, 2023, doi: 10.1109/ACCESS.2023.3261247.

9. H. L. Gururaj, B. C. Soundarya, S. Priya, J. Shreyas and F. Flammini, "A Comprehensive Review of Face Recognition Techniques, Trends, and Challenges," in IEEE Access, vol. 12, pp. 107903- 107926, 2024, doi: 10.1109/ACCESS.2024.3424933.

10. K. Man, J. Chahl, W. Mayer and A. Perera, "The Effects of Different Image Parameters on Human Action Recognition Models Trained on Real and Synthetic Image Data," in IEEE Access, vol. 12, pp. 95223-95244, 2024, doi: 10.1109/ACCESS.2024.3423344.

11. T. Gim and K. -A. Sohn, "Regularization Using Noise Samples Identified by the Feature Norm for Face Recognition," in IEEE Acess, vol 12, pp. 123267-123275, 2024, doi: 10.1109/ACCESS.2024.3453030. for Face Recognition," in IEEE Access, vol. 12, pp. 123267- 123275, 2024, doi: 10.1109/ACCESS.2024.3453030.

12. B. -S. Kim and S. Seo, "Intelligent Digital Human Agent Service with Deep Learning Based- Face Recognition," in IEEE Access, vol. 10, pp. 72794-72805, 2022, doi: 10.1109/ACCESS.2022.3188852