WWW.IJRPR.COM

International Journal of Research Publication and Reviews

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

Intelligent Video Retrieval Technology and Cascaded Intelligent Face Detection Algorithm in Video Monitoring Systems

G. Jaswanth Kumar

GMR Institute of Technology (GMRIT) Srikakulam

ABSTRACT

As security needs increase, it is important to integrate intelligence into video surveillance. This research provides an integrated framework for the development of security surveillance that simultaneously combines face detection, recognition, and video surveillance. The main goal is a state-of-the-art multi-face detection algorithm that forms a solid foundation suitable for video surveillance. Advanced facial recognition improves this by improving situational response by providing real-time visual performance. The framework's new AI-powered video retrieval and alignment approach enables fast and accurate analysis of key events. To solve the low resolution issue, the use of reinforcement learning is used to improve content-based video frame retrieval to ensure the performance of the system across different videos. This is changing public access security systems.

Key words: *Video processing, face recognition, bracketing function, intelligent video acquisition, video monitoring system.*

1. INTRODUCTION

Video surveillance systems have become an important part of security systems. Modern security and surveillance infrastructure supports many applications such as law enforcement, public security and private security. Intelligent video access technology represents a revolution in the way video data is analyzed and used in video surveillance. This technology uses various computer vision and artificial intelligence algorithms to identify, mark and store specific events or objects in a video stream. It can identify vulnerabilities, track objects of interest, and identify patterns; This makes it useful for security experts, researchers and analysts. Face detection plays an important role in identifying people, tracking their movements and analyzing their behavior.

The progressive intelligent face detection algorithm is an advanced face detection method designed to increase the accuracy and speed of face recognition in complex video environments. The algorithm works by using a series of cascading stages that refine and narrow the search for faces in the video at each stage. The combination of intelligent video acquisition technology and progressive intelligent face detection algorithm opens up new possibilities for video surveillance.

2. LITERATURE SURVEY

2.1. Dong, Z., Wei, J., Chen, X., & Zheng, P. (2020). Face detection in security monitoring based on artificial intelligence video retrieval technology. Ieee Access, 8, 63421-63433.

The purpose of this paper is to propose a video-oriented progressive intelligent face detection algorithm using deep learning and face detection neural network to ensure the face is clear and accurate. The algorithm used in this article is a cascading intelligent face detection algorithm that builds a deep learning network by cascading various features, including edge features, contour features, local features, and semantic features. Advantages of this algorithm; good detection results, good performance for single-face and multi-face images, strong robustness for face rotation, fast speed, and real-time detection of face may be needed. Simulation results show that the algorithm has good detection performance and accuracy in various conditions. The disadvantage of this form is that it does not return to true when objects or faces are occluded.

2.2. Badave, H., & Kuber, M. (2021, March). Face recognition based activity detection for security application. In 2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS) (pp. 487-491). IEEE.

The purpose of this document is to propose a workaround to implement facial recognition in security applications. The algorithm used in this article involves generating encodings for each face based on simple measurements and comparing them to known faces in the training data. The advantages of this approach include the integration of person identification with the division of labor, which can be used for many purposes, such as preventive use.

The results showed that the method performed with good accuracy on four different tasks. Limitations of facial recognition such as transitions, occlusions, and lighting changes can be addressed in future studies.

2.3. Jo, W., Lim, G., Hwang, Y., Lee, G., Kim, J., Yun, J., ... & Choi, Y. (2023). Simultaneous Video Retrieval and Alignment. IEEE Access, 11, 28466-28478.

The purpose of this article is to propose a new task called simultaneous video retrieval and alignment, which meets the need to search for similar videos (video retrieval) and align the position of pair-related videos. The algorithm used in this article is the Simultaneous Video Acquisition and Alignment Framework (SRA), which is a two-stage process with a front-end proposal phase and a downstream phase. Advantages of the SRA framework include the ability to deal with poor video quality, support for relational video-level and hierarchical segment-level annotations, and video ingestion and operation in an integrated manner during decision-making. The results of the experiment in the paper show how the SRA framework works simultaneously to enable video playback and complete the task. A new metric called Jaccard weighted average sensitivity (mAPJ) has been proposed to evaluate the performance of this task. The limitation of this approach is that it requires a lot of training data. Moreover, the plan can only create physical harmony between similar films.

2.4. Ullah, T., Khan, A., & Waleed, M. (2020, June). Reinforcement Learning for Enhanced Content Based Video Frame Retrieval System in Low Resolution Videos. In 2020 12th International Conference on Electronics, Computers and Artificial Intelligence (ECAI) (pp. 1-6). IEEE.

The purpose of this article is to utilize additional information to improve the content-based video framing system for low-resolution videos. The algorithm used in this article is a combination of fast feature extraction (SURF) for inference and reinforcement learning (RL) for point of interest (POI) calculation and reduction pole. The advantages of the proposed method include the ability to remove unnecessary frames from the video and store similar frames as query images using content based on the Image Retrieval concept (CBIR). The results demonstrate the effectiveness of the system and provide a benchmarking platform. A limitation of this method is that it requires a lot of training data to train the DQN. Moreover, the proposed method can only improve the performance of CBVFR systems in low-resolution video.

2.5. Guo, Q., Wang, Z., Wang, C., & Cui, D. (2020, July). Multi-face detection algorithm suitable for video surveillance. In 2020 International Conference on Computer Vision, Image and Deep Learning (CVIDL) (pp. 27-33). IEEE.).

The algorithm used in this article is an improvement of the MTCNN (Multi-Task Cascaded Convolutional Network) algorithm, which combines data reliability, depth separation convolution, and optimized NMS(not maximum) algorithm rule. The advantages of this algorithm include faster processing, faster detection of many faces in crowded environments, and greater detection. Research results show that the developed MTCNN algorithm achieved 92% accuracy in detecting multiple faces. The proposed algorithm can only detect faces in the front view. This is the limitation of video surveillance application which can capture the face from different angles.

3. METHODOLOGY

3.1. BASE PAPER: Face detection in security monitoring based on artificial intelligence video retrieval technology.

Developing training and validation: The model is trained using the CASIA webface database, which includes a variety of people of different races and ages. The training process is balanced, there are more than 100 people in each class and a total of 915 people are selected. 80 samples for each user are used for training and another 20 samples for validation..

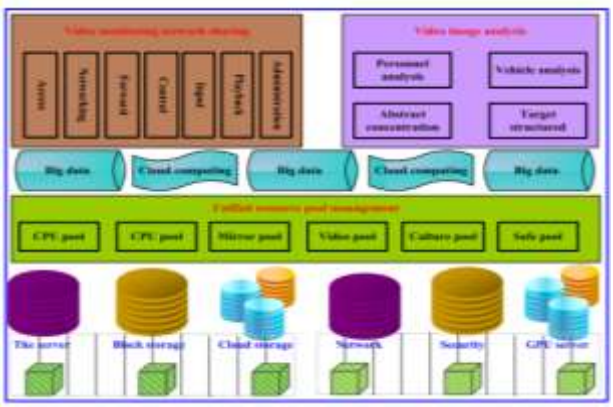


Fig 1. Intelligent video monitoring architecture.

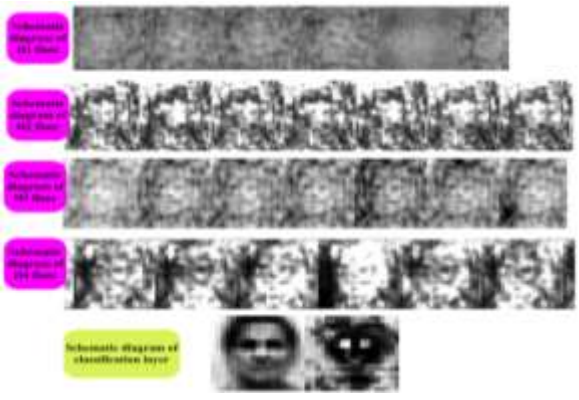


Fig 2. Features of hidden layer and classification layer after optimization.

Face preprocessing: This includes three steps: face correction, data enhancement and normalization. Face correction uses affine transformation to correct distorted faces. Data augmentation involves working up and down to improve the model's capabilities.

Feature extraction and training: The model is trained using the network model and the output feature vector is extracted by the feature extraction part. Training involves backpropagation to update the parameters of the multilayer network layer by layer. The aim is to make predictions based on the given results..

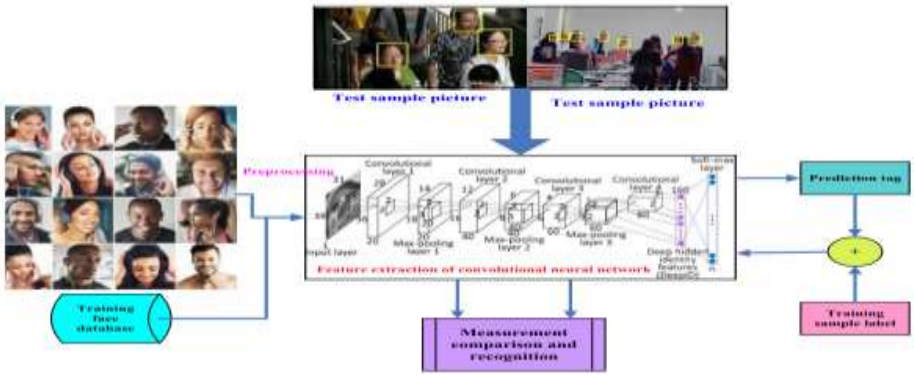


Fig 3. Recognition process based on deep convolution neural network

Combining full and local feature networks: The proposed model combines full feature network and local feature network, which is more compatible with the physiological process of the human body. The entire face in the sample set is used as the total facial feature, and local features are extracted based on the landmark content on the face.

Correlation analysis: Perform correlation analysis after getting all lighting and local settings. The total set includes 915 species of humans, with 100 examples for each group. The localization process involves different local features of human faces that are extracted as unique characters of the face.

The performance of this model is calculated by network model training, feature extraction, Linkage.relate global and local features and based on extracted features. This model aims to improve face detection and recognition performance, as shown by the experimental results in the article.

3.2. Face recognition based activity detection for security application.

There are two main methods of paper design: face recognition and recognition function.

Face Recognition: Live Video Capture and Face Detector: Capture input image from camera using IP network camera. Convert the image to grayscale and then get the HOG representation to see the face in the image.

Find face pose: Find face orientation by warping each image so that lips and eyes are in the same place in the image using face landmark algorithms.

Create coding base measurements for each face: Extract base measurements to find a known face with closest measurements, including parameters such as ear size, nose length, and eyes-to-eye distance.

Face detection and labeling with coding: It compares the images of known people in the training data with the test images, and if similar facial patterns are available for sale, the person is recorded and identified**.**Find the face and label the face from encodings: The image of a known person present in the trained dataset is compared with the test image, and if a similar pattern of the face is found, the person is labeled and recognized.

Work Experience: Real-time Pose Estimation with OpenPose: OpenPose enables real-time capture of highlights of multiple people from a single image including hands, feet, body and face Important.

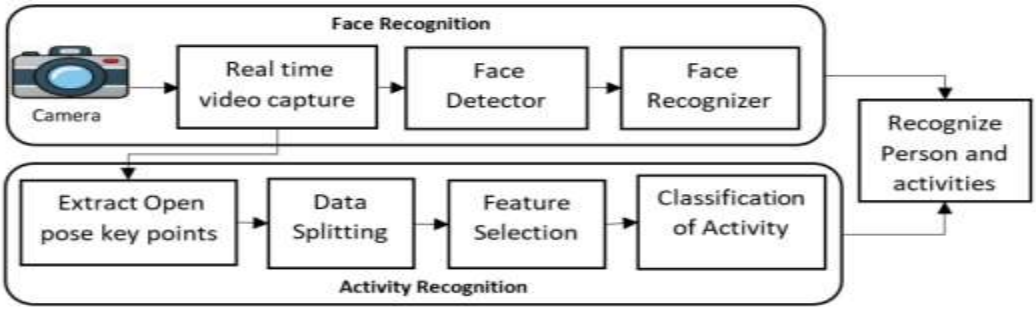


Fig. 1. Proposed Architecture

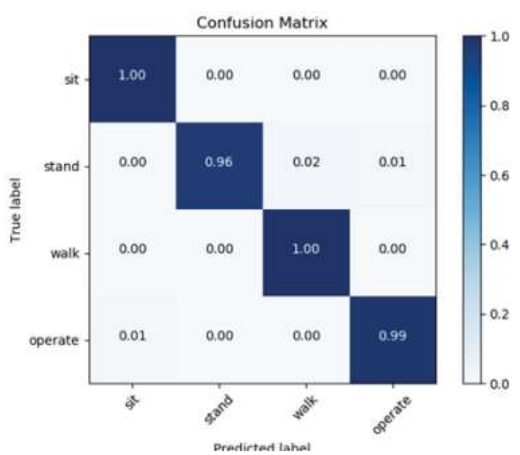


Fig 2. Confusion Mat

No.	Activity	Recognition of Activity and Person	
1	Sit		
2	Stand		
3	Walk		
4	Operate		

Table1. Recognition And Activity Classification

It has three modules: hand detection, face detection and body + foot detection. An RGBimage is considered as input and the design of a whole-body pose estimation network incorporating Part Affinity Fields is used.

Functional design includes real-time video capture, face detection, predictive modeling and knowing how to use OpenPose. The facial recognition method involves capturing live images, finding facial expressions, generating numbers for each face, and then tagging the faces by number.On the other hand, the recognition function uses OpenPose to estimate human poses and identify various key points to identify different tasks.

3.3. Simultaneous Video Retrieval and Alignment

The Synchronized Video Acquisition and Alignment (SRA) model provides new functions and methods for synchronized video acquisition and alignment. The SRA framework has several key components, including advanced recognition, video-level inference, and similarity learning.

Foreground Bidding Phase: The foreground bidding phase is designed to eliminate the unaffected ambiguous or irrelevant part of the film and determine the value area by looking for a transition shot or a sequence containing transitions or transitions. expectations. Less relevant to video level content. This phase involves creating suggestions from the relationship between cross-sectional characteristics and calculate the confidence of each suggestion, in order to first use information about the situation to represent the probability of a particular event. event plan.

Fig 1. . Composition example of an untrimmed video related to aforest fire

Video level feature extraction: video level features Extraction involves using the TimeSformer network to extract average video-level features from negative film samples. These average features are calculated for each triple frame and used to represent the pattern in the long-term video display.

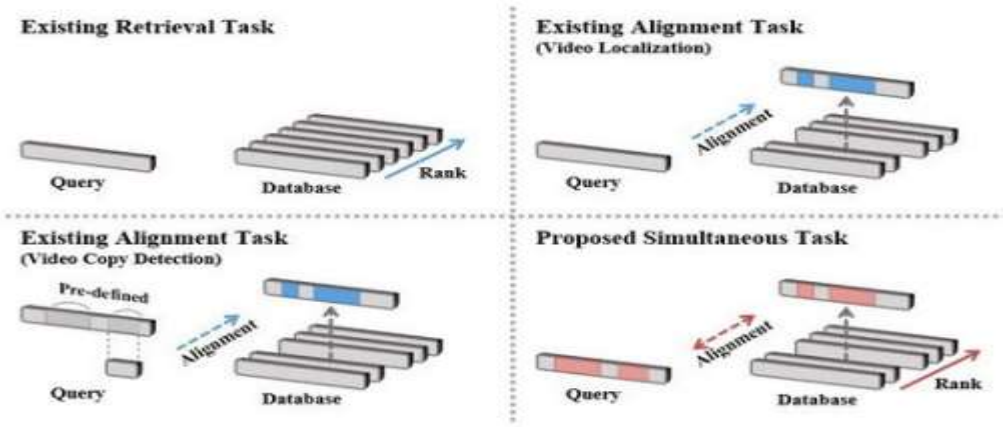


Fig 2. Comparison of tasks related to video retrieval and alignment

Similarity learning: Similarity learning uses metric learning based on similarity maps of adjacent images to learn distance based on proximity, and also focuses on the long-term ability of the model to represent interior space.

In summary, the SRA framework addresses the limitations of people who already have a way to get the job done independently by integrating foreground bidding, video-level feature extraction, and similarity learning to perform simultaneous video retrieval and alignment

3.4. Reinforcement Learning for Enhanced Content Based Video Frame Retrieval System in Low Resolution Videos.

The working concept has two main parts: removing unnecessary frames and restoring similar video frames. The content-based video framing system (CBVFR) framework has the following methods:

Redundant removal:Convert user input video into frames. Use HOG feature extraction to remove excess from frames and convert them into smaller patches.

Feature Extraction and Quantization: Extract SURF-based features from the training dataset using bag of features (BoF) extraction technique. These features are quantified using k-means clustering technique to reduce feature vectors. A histogram of comments is created by applying the coding method to the results of the comments. Practical method.

Fig 1. Redundant video frames removal using HOG features and Euclidean similarity techniques

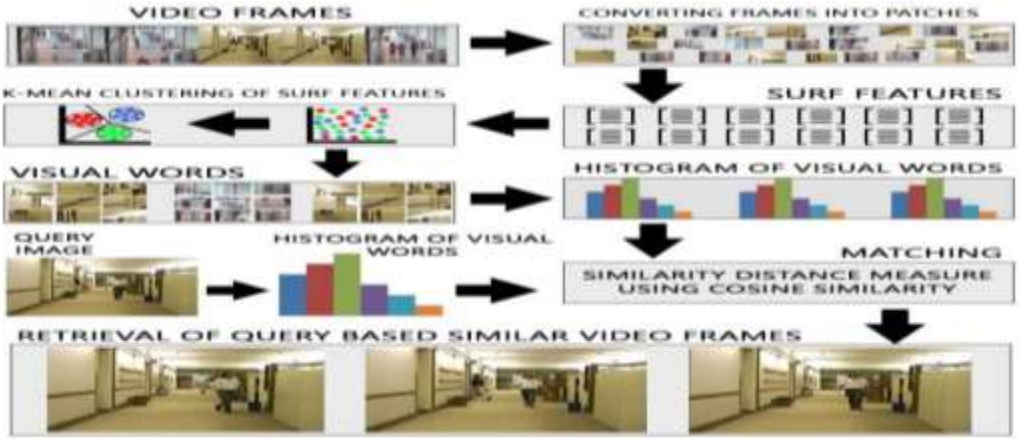


Fig 2. Proposed framework for the retrieval of similar video frames from a video.

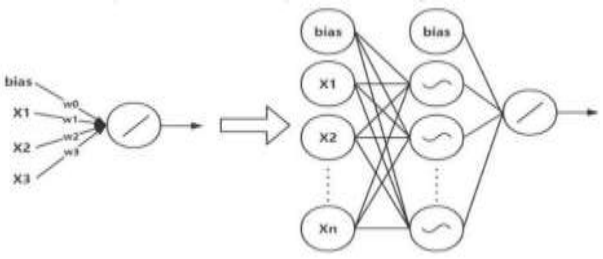
Similarity Measurement and Retrieval:From Users Image or features are extracted from image content using the same techniques as video as squares. Cosine similarity is used to measure the distance similarity between the query image and the corresponding video. Then add as many frames as you want. The model works by first removing redundancy from video frames using HOG feature extraction and then extracting SURF-based features from the training data.

Features were calculated using k-means clustering to reduce the size of vectors. The coding process is used to create histograms of observed results. When the user requests an image, the same technology as the video frame is used to extract features. Use cosine similarity to measure the distance similarity between the query image and the non-repetitive video and measure the number of frames to be returned by the method.

3.5. Multi-face detection algorithm suitable for video surveillance.

Entries in the article include the following:Averaging function: The article uses a convolutional neural network (CNN) to learn weight results from the Training sample record. The model aims to optimize the mean by passing the gradient of the loss function back through the layer.

P-Net development: The study aims to reduce the number of competitors by improving the quality of the P-Net module. This is done by enhancing the results of face quality analysis providing more detection information for subsequent face recognition. The proposed model aims to improve P-Net to generate fewer candidate faces, thus making the entire discovery process faster.



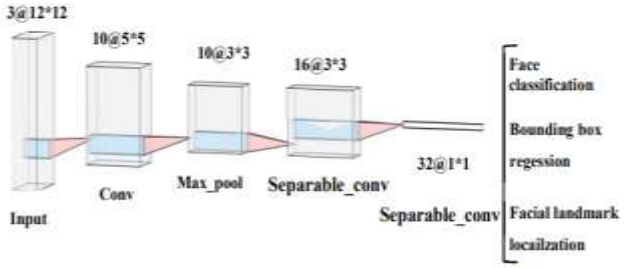


Fig 1. Convolutional Neural Network.

Fig 2. Convolutional Neural Network.

Discrete correlation: A report on the relationship between correlation and spatial correlation of convolutional layer channels; can be shared together for better results.This method aims to improve the performance of the model by optimizing the correlation and spatial correlation of the convolution layer.CNN training will be used to collect samples in order to learn the weights in the operation of the proposed model.The gradient of the loss function is then passed back layer by layer to optimize the averaging process. Additionally, the P-Net module has been improved to improve the quality of candidate faces and reduce the number of candidate faces, thus accelerating the entire detection process. In addition, the separation of correlation and correlation of the convolutional process is used to obtain a better performance model.

Fig 3. NMS Removal of redundant candidates.

4. RESULTS AND DISCUSSION

4.1 Face detection in security monitoring based on artificial intelligence video retrieval technology.

The model proposed in the article combines all features with a central feature network for face detection and recognition. Methods described in this article include using the CASIA online face database to develop training and validation, face-first procedures for facial treatment, developing data and using network models for normalization, extraction and training, and establishing allconnections. and local. Participate in feature networks and recognition-based feature extraction.

The operation of this model involves training the network model, extracting features, combining global features with local features, and realizing relationships while drawing conclusions. As the test results given in the article show, this model is designed to improve the artifact's ability to see and recognize faces. Experimental results show that the proposed model outperforms existing models in terms of actual quality, negative cost and overall performance, as shown in the ROC curve comparison and performance comparison table. The combination of several models has proven to be better than using a single model, and the proposed model has shown good performance in face detection and recognition.

Overall, the results and discussion of this paper demonstrate the effectiveness of the design in improving face,recognition in security assessment application. With implications for the development of video surveillance systems and biometric recognition technology

4.2. Face recognition based activity detection for security application.

The design model offers a new way of combining recognition of people with their work, resulting in greater accuracy in identifying tasks and participants.

The fact that training and usage is reduced to almost zero after 60 times shows the effectiveness of the model and training materials model in the study. The combination of face recognition and task recognition yielded good results, with the model accurately classifying the activities a person was doing and identifying participants in water.

The model was tested on four different tasks performed by four different people andthe tasks and participants were identified, revealing its potential for a real-world application form.

The proposed facial and facial recognition technology has many potential applications, including medical, preventative, security, and surveillance. The performance of this model can be extended to larger projects, personal data and outdoor projects, improving its applicability in various situations.

Using OpenPose for instantaneous prediction and combining it with activity recognition of faces provides a powerful framework for recognizing and classifying human activities, making Model planning useful for many practical applications. In summary, the proposed combination model for face recognition and task classification has shown good results with applications in many fields. The accuracy, robustness, and measurement capabilities of this model make it suitable for use in different real-world situations. Table 1. Training and Simulation Parameters

4.3. Simultaneous Video Retrieval and Alignment.

Simultaneous Video Acquisition and Alignment (SRA) model has proven its effectiveness through many tests and reviews. The SRA framework includes progressive mapping, video extraction, and similar learning and is evaluated using the FIVR-5K dataset and FIVR+A-5K dataset.

Visualization of the results of video-level features shows that SRA with forward-looking statements exhibits significant discriminatory power compared to other state-of-the-art methods. Also showing the ability to recover time-lapse video, SRA works well for longer videos thanks to the architecture that does not require integration for the process in the front-end and ingest operations.

The efficiency and trade-off of SRA were analyzed and the results show that SRA has the highest level of improvement compared to other methods, although the hypothesis is different.

The foreground concept level is introduced to improve the retrieval ability, effectivelyremoving blurry or irrelevant areas from the negative image as desired.Additionally, the SRA model is evaluated using the FIVR + A dataset and multivariate analysis comparing the results with other methods.

The results demonstrate the effectiveness of SRA in working to recover and complete tasks.The report, results and discussion demonstrate the effectiveness and efficiency of the SRA model in solving video retrieval and completing the task at the same problem time and demonstrate its robustness with Distinctive features. capability, efficiency and performance compared to current state-of-the-art methods.

4.4. Reinforcement Learning for Enhanced Content Based Video Frame Retrieval System in Low Resolution Videos.

The proposed content-based video frame retrieval system (CBVFR) framework involves removing duplicate frames and retrieving similar images. This study concluded that CBIR and CBVFR have valid information in real-time monitoring and data collection and can be used for video content. The plan can reduce storage space by removing duplicate frames and video content becomes better by recovering necessary frames from the entire video.

Studies have also addressed the use of algorithmic techniques to extract features from video frames and the use of similar methods to achieve better selection and analysis.The proposed model works by first removing duplicates in the image using HOG feature extraction and then extracting SURF based on the features of the training data. This feature involves the use of k-means clustering to reduce the size of image vectors. Use coding techniques to create histograms of observed results. When the user requests an image, the same technology as the video frame is used to extract features. Use cosine similarity to measure the distance similarity between the query image and the non-reciprocal video and store the required number according to the similarity measure.

This study also addresses the use of SURF feature extraction in the proposed system as it is invariant with scaling, rotation, translation and illumination. Even if the question images are in different poses, the system can take the same frame. The proposed model achieves high accuracy in video classification, and the calculation time of each frame varies depending on the video type.

In summary, the CBVFR program effectively solves the problem of removing repeated frames and restoring similar images, as well as practical applications in analysis, information in real-time storage,and video summarization. The potential of the proposed system is further strengthened by the use of standard techniques for feature extraction, selection and recognition, as well as the potential for future research using convolutional neural networks (CNN), deep learning and CUDA models

4.5. Multi-face detection algorithm suitable for video surveillance.

The model proposed in the article is successful in detecting faces accurately and quickly.Minimization of the algorithm involves using CNN training collection models to optimize execution. Improvements were made to P-Net to improve quality and reduce the number of facial candidates, leading to increased detection speed and accuracy. Additionally, a combination of correlation and convolutional methods is used to improve the performance of the model.Experimental results show that this algorithm is better than the MTCNN model in terms of accuracy and It's especially fast in many face detection scenarios. The improved P-Net module demonstrates improved face recognition accuracy and fast time, with accuracy up to 90.00% and average detection speed up to 18.6%.Overall, the model shows great promise in improving the accuracy and speed of face detection, especially in cases where there is more than one face. The development of the P-Net module and optimization of performance contribute to the improvement of the performance model.

Algorithm	**Accuracy**
Cascaded face detection	82%
SRA	85.3%
SURF & RL	84%
Activity based detection	82%
MTCNN	92%

5. CONCLUSION

This paper presents a new method to combine facial recognition with human task classification using camera-based non-contact protection. The system is recognized by many people working in a variety of indoor environments and shows great accuracy in identifying people and the work they do. The plan can be used in prevention, safety, health care and performance evaluation. The article also suggests future extensions and addresses the limitations of facial recognition. Additionally, the proposed method is used to support learning and content-based image retrieval to improve video frame retrieval in low-resolution videos, and its results are seen through simulations of sample datasets and self-shot videos. In addition, this article discusses the techniques used for 3D face reconstruction using RGB-D cameras, regarding the impact of distance on 3D reconstruction and the use of multidimensional kernel filters to improve the editing of the representation of 3D objects. While the multi-face detection algorithm suitable for the proposed video analysis performs better, the simultaneous video retrieval and alignment (SRA) method also provides good results in video retrieval and performance

6. REFERENCES

[1].Dong, Z., Wei, J., Chen, X., & Zheng, P. (2020). Face detection in security monitoring based on artificial intelligence video retrieval technology. Ieee Access, 8, 63421-63433

[2].Badave, H., & Kuber, M. (2021, March). Face recognition based activity detection for security application. In 2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS) (pp. 487-491). IEEE.

[3].Jo, W., Lim, G., Hwang, Y., Lee, G., Kim, J., Yun, J., ... & Choi, Y. (2023). Simultaneous Video Retrieval and Alignment. IEEE Access, 11, 28466-28478.

[4].Ullah, T., Khan, A., & Waleed, M. (2020, June). Reinforcement Learning for Enhanced Content Based Video Frame Retrieval System in Low Resolution Videos. In 2020 12th International Conference on Electronics, Computers and Artificial Intelligence (ECAI) (pp. 1-6). IEEE.

[5].Guo, Q., Wang, Z., Wang, C., & Cui, D. (2020, July). Multi-face detection algorithm suitable for video surveillance. In 2020 International Conference on Computer Vision, Image and Deep Learning (CVIDL) (pp. 27-33). IEEE.