Abstract

Face recognition has evolved in various ways over the past three decades. Face recognition research has traditionally concentrated on recognising faces from still images. The focus of study has recently shifted increasingly towards video-based approaches. Face films may now be captured, stored, and analysed thanks to the development of low-cost video cameras and enhanced computational power. People generally display a lot of facial position and expression fluctuations in video; therefore, video inputs provide rich and redundant information in the form of several frames.

A video index that describes the video content is required to browse, search, and work with video documents. It serves as the basis for applications such as multimedia-rich digital libraries and filtering algorithms, which automatically search for relevant video documents based on user profiles. The indexes need to be as complete and extensive as feasible to accommodate these diverse applications. Until now, documentarists have mostly created indexes by hand, assigning a limited number of keywords to each video. Due to the specialized nature of the work, manually indexing video documents is time-consuming and expensive. Automatic classification of video content is therefore required.

The technologies that enable people to record and distribute digital video data easily are advancing rapidly and becoming increasingly accessible today. While high-speed and dependable networking has shifted towards mobile and wireless access, personal computers continue to become faster, smaller, and less expensive. The Internet and portable devices are now widely used for creating and sharing video documents. Because video contains a wealth of content for numerous important applications, its use as one of the most popular media formats has increased significantly. Further improvements are needed on existing Content-Based Video Retrieval (CBVR) systems to support and sustain the expansion of video content.

The human face is one of the most crucial visual elements in a content-based video retrieval system. Since face detection and tracking from video automatically identify and locate the face region in input frames, they are essential parts of face recognition systems. Typically, the face recognizer receives the necessary features from the located face region. A brief overview of research on video indexing and retrieval using human faces as cues is presented in this thesis.

With applications ranging from biometric identification to video search and retrieval, visual surveillance, and human-computer interaction, face recognition in video has generated significant interest over the last decade. Despite significant advancements, the issue remains challenging to resolve for videos captured in unrestricted settings. A further research study discusses the need for "Video indexing and retrieval taking human faces as cues" to address various problems. This work improves the following aspects: (1) increases the size of the training data set to improve recognition accuracy; (2) uses new deep and machine learning algorithms to improve the accuracy of face detection;(3) aims to address significant challenges such changing position, having a limited amount of storage space, losing crucial video frames, and the intricacy of time and place; and (4) look into ways to use a faster and better face recognition technique to increase retrieval speed and accuracy. This effort will address the primary challenges outlined in several studies, including the use of databases, the quality of video data, and computational costs.

Video processing technology is receiving more attention to reduce network transfer stress, analyze video data efficiently, and extract reliable information from videos based on human faces. Key frame extraction and video segmentation significantly reduce the data used in video processing. "Video indexing and retrieval using the human face as a cue" has consequently gained more attention in research. This research aims to identify unlawful activity in the military or armed forces utilizing video surveillance systems. Aside from that, this discovery is crucial for various video face identification applications. The aforementioned study is necessary to enhance effective communication and communication speed in a communication channel. This thesis's primary goal is to develop a video indexing and retrieval technique that utilizes a person's face as a cue to aid in identifying them across various video clips. The work enhances the following features:

(a) Major issues, including posture changes, storage capacity limitations, the inability to save crucial video frames, and the complexity of time and space, are all addressed in this thesis.

(b) To improve recognition accuracy, the size of the training data collection must be increased.

(c) To increase the precision of facial detection.

(d) Examine how to use a faster and more effective face recognition technique to increase retrieval speed and accuracy.

Face detection in videos is a challenging task, as evidenced by the analysis of existing techniques for indexing videos using human faces. Angular variations of the face, changes in posture, illumination-invariant aspects, time and spatial complexity, and the inability to save key frames from videos are the primary sources of this issue. This thesis presents a novel approach to video indexing that utilizes various machine learning and deep learning algorithms, along with human faces as cues, to address these challenges. Using a variety of machine learning and deep learning techniques along with human faces as cues, this thesis explored a novel approach to video indexing that addresses these issues.

(a) The initial stage in indexing a video using a human face as a cue is pre-processing the incoming video by removing frames. Faces are recognized as frames from the input video.
(b) Faces identified by machine learning or deep learning algorithms are subjected to features (key frames).
(c) Machine learning or deep learning algorithms are used to recognize or classify faces.
(d) Faces are then indexed as standard faces or unique barcodes and stored in a database.
(e) Then, when required, obtain the index data from the Query System.
(f) The output can be bar codes of any type or photos of faces.

This dissertation's primary goal is to utilize machine learning and deep learning methods to develop automated video indexing and retrieval systems that leverage human faces as cues. They employ Viola Jones, MTCNN, DSFD, Blaze-face, YOLO v3, YOLOv5s, and YOLOv8n, among others, for face detection and frame extraction from the input video. The color histogram approach, combined with various machine learning and deep learning algorithms, extracts key frames from a sequence of images. These essential frames are then stored in a database as a standard human face and a linear EAN-8 bar code or QR code to index and retrieve videos. Therefore, the primary contribution of this dissertation can be summarized as follows.

- **Pre-processing and frame extraction from the input video** - Dynamic video combines the frame and the recorded scene. Thus, extracting still images from input videos, represented as scenes, shots, and pictures, is the first stage. A sequence of filmed and shot frames makes up the scene. Instead of extracting frames from the input video for the experiment, direct frames are taken from several video datasets.

- **Key frame extraction from the frame-** The key frame captures the essential details of each shot. The key frames in this study feature human faces with unique expressions, postures, lighting conditions, and illumination. In this dissertation, the key frame is extracted from each frame of a particular video using the Colour Histogram approach.

- **Face image cropped from the key frame** - Several machine learning and deep learning algorithms (Viola-Jones, MTCNN, Shuffle Net, Combined MTCNN and ShuffleNet, DSFD, Blaze Face, YOLOv3, YOLOv5s, and YOLOv8n) are used to detect faces from the extracted key frames. Human faces with distinctive expressions, age, postures, lighting circumstances, and illumination directional changes can be clipped from key frames using these machine and deep learning algorithms. The "Hollywood Video Data Set, YouTube Face Video Dataset, TV Series Video Database, Face Video Dataset FDDB, Face Video Dataset WIDER, Face Video Dataset LFW, Face94," "YaleB Face Database," "Face database FERET," and FG-NET dataset, among others, are used to test this approach. To determine the gradient of the face image, the faces from key frames must be converted to greyscale. Image gradients are computed from this greyscale image (faces) using distinct techniques, including LGFA and the sliding window approach.

- Using human faces as cues to index the incoming video, a unique QR code, a linear EAN-8 bar code, and identification from the key frame of the input video are created and saved as an index for each face

- Remove the index data from the query system. Next, retrieve the index data from the Query System as needed.

The dissertation begins with a brief overview of video representation, advanced methods for indexing and retrieving videos from video surveillance systems that use human faces as signals,

segmenting video documents for indexing purposes, indexing videos from video documents using a variety of modalities, the importance of using human faces as cues in video indexing, and other topics. To analyze the video resources, video document segmentation is necessary for indexing purposes. This thesis covers several state-of-the-art methods for indexing and retrieving videos from video surveillance systems that utilize human faces as a signal.

In this dissertation, a novel framework has been presented for video indexing through face images using Barcodes. A survey and analysis of several studies indicate that person detection is challenging when video indexing is performed using low-level attributes. A person's facial expression, posture, mood, lighting variations, and occlusions are all important factors in video indexing using the human face. Moreover, the intrinsic uncertainties of video-based identification were observed, including changes in location, sensitivity to low resolution, and partial occlusion of facial features. From the perspective of indexing and storage space, all of these techniques are time- and space-intensive. We created a video indexing method called "Video indexing through human Face Images" to overcome all of these problems. To identify faces in the video, this suggested method of video indexing by face recognition utilizes an EAN 8 linear bar code. This technique employs the color histogram method for key frame detection and the Viola-Jones object detector for face detection. The EAN 8 bar code is used to index the face as a bar code, and the sliding window method is used to determine the picture gradient. This approach also considers occlusion, illumination variation, facial expressions, and minor shifts in face direction while generating an EAN 8 barcode from human faces in videos.

The proposed procedure for generating linear EAN-8 barcodes from facial images is based on the combined window and LGFA technique, gradient calculations, gradient directions, normalization, and quantization. The final step is to convert input faces to tags using EAN-8 linear standardized identifications.

In video indexing and retrieval, deep learning and image processing are utilized to increase storage capacity by saving important frames from the movie and to decrease the time and space complexity of extracting angular faces from videos. Thus, to tackle these issues, a novel hybrid sliding window and LGFA technique, the EAN 8 barcode for use as a face index, and the Viola-Jones Algorithm for use as a face detector.

In image processing for video indexing and retrieval, deep learning and machine learning are utilized to increase storage capacity for key movie frames and reduce the time and space required to learn angular faces from videos. We discuss the use of the linear EAN-8 bar code as a face index. To get over these issues, the MTCNN Algorithm has been proposed for use as a face detector. The scenario of illumination of the invariant facial picture is also created using this format, and the computation is straightforward. EAN-8 linear barcodes finally indicate input faces. While decreasing the complexity of time and space, the proposed strategy enhanced storage capacity. This improves the video indexing and retrieval methods in the recommended manner.

A novel method is proposed for Video indexing through face images using QR codes. In image processing for video indexing and retrieval, deep learning enhances the storage capacity for key frames from videos and reduces the time and space complexity of extracting angular faces from videos. To address these issues, the MTCNN Algorithm has been proposed as a face detector, and the QR code is utilized as a barcode to facilitate face indexing. Additionally, this form is used to generate the lighting situation of the invariant face picture, which is computationally simple. The suggested approach increased storage capacity while reducing time and space complexity. One benefit of a QR code is that it will continue to function correctly even if part of it is damaged or incomplete. A QR code is created once the input image (a face) has been scanned both horizontally and vertically.

The primary function of the MTCNN algorithm in this method is to recognize and align faces in angular key frames. The suggested method significantly enhances video indexing and retrieval techniques.

We presented two innovative techniques for video indexing using deep learning models and face images. The first method is "Video Indexing through Human Faces by Combined Deep Learning Neural Networks," which combines the Shuffle Net and MTCNN algorithms for face detection from the input video. In this section, the accuracy of face detection is enhanced by combining Shuffle Net and MTCNN, while the accuracy of face recognition is improved using Eigen faces. Initially, faces are detected by combining the Shuffle Net and MTCNN techniques, followed by face identification using the Eigen face approach and principal component analysis (PCA). Consequently, the work demonstrates how the two components can be integrated to produce a thorough, reliable, real-time face detection solution. The results indicate that the

proposed system can recognize individuals in various situations and in real-time, with a face recognition learning rate of 99.35%. The results are encouraging because the proposed method can be implemented on a device due to its short processing time.

The YOLOv5s and YOLOv8n algorithms for face detection from the input video, are used in the second suggested method, "Video indexing and retrieval through the human face as a cue using Light Weight Deep Learning algorithm YOLO v5 and YOLO v8." to index videos by face identification. In image processing, lightweight deep learning is utilized for video indexing and retrieval to enhance storage capacity for important video frames and reduce the time and space required to capture angular faces from videos. It facilitates facial recognition in portable devices. A face detector based on the YOLOv8n algorithm has been developed to address these problems. For video indexing, the human face is utilized.

Using video indexing technology, an account's personal search, authentication, and affirmation can be specified. Applications for this technology include communication channel description, video surveillance, security, and human activity detection.