International Interdisciplinary Virtual Conference on
'Recent Advancements in Computer Science, Management and Information Technology'
International Journal of Scientific Research in Computer Science,
Engineering and Information Technology | ISSN : 2456-3307 (www.ijsrcseit.com)

A Review on Face Recognition in Video Images

Gajanan Joshi, Dr. Prakash Khanale
Department of Computer Science, Research student, S.R.T.M.U., Nanded, Maharashtra, India

ABSTRACT

There has been a lot of study done on face recognition using images of faces, but there hasn't been nearly as much done on video-based face recognition. Recently, scientists have dedicated a lot of time and energy to solving the challenge of facial recognition. The purpose of this study is to present an up-to-date overview of key research that has been done on human face recognition using video images. In the first part of this article, we will provide an overview of face recognition and the uses it has. The next step is a discussion of a literature study on the most recent developments in face recognition algorithms. Although many still pictures or video sequences may be degenerately treated as a single still image, they really possess a number of unique properties of their own. Three characteristics that may be seen in a series of still photographs or videos are summarized in this chapter. Then, using the exhibited qualities, we discuss newly suggested methods such as Eigen-faces, Neural Networks, Fisher Discriminant Analysis (FDA), Artificial Neural Network Fuzzy classification, and Genetic algorithm.

Keywords— Face recognition, Video images, soft computing, PCA, LDA, FDA, Adaptive Skin color model, Eigen-faces. Artificial Neural Network, Fuzzy Logic and Genitic Algorithm.

I. INTRODUCTION

Face recognition is a significant research issue that crosses many professions and disciplines. This is because of the many practical applications of facial recognition technology, such as in ID verification, criminal background checks, bank card verification, security and surveillance, and mug shot searches. It's a trait shared by all humans and crucial for effective communication and social engagement. As video recording technology has become more commonplace, researchers have begun to concentrate their efforts on video-based scenarios in which facial recognition may be used. In video-based face recognition, the video sequence of an individual is used as the query or target in order to identify that person. There are primarily three instances when video-based facial recognition is useful. An example of this is the V2S (Video-to-still) scenario, which evaluates a user-provided video against a database of still images of a person's face (mug shots, ID pictures, licence photos, etc.), These are often consumed in well regulated settings, and as a result are of superior quality. Systems for screening watch lists often need to deal with this circumstance. Instead, In the S2V (Still-to-Video) scenario, In order to find a person of interest in recorded surveillance footage, the application compares a still face snapshot to a database containing video sequences. The third and final scenario is referred to as the "V2V (Video-to-

Video) case," and it involves comparing a single video sequence to a group of target video sequences. Such a system may be put to use, for instance, to locate a missing person by comparing and contrasting images from several surveillance cameras.

II. LITERATURE REVIEW

Citation no.	Author	Methodology	Results	Limitations
[1]	Zhiwu Huang, Shiguang Shan,	a benchmarking and comparing research that was based on the freshly compiled COX1 Face DB still/video face database viewed and experimentally contrasted several set-based techniques in use included a brand-new Point-to-Set Correlation Learning (PSCL) technique.	Their face recognition approaches for V2S/S2V tasks may also be used to the V2V task, and they ultimately get encouraging results for this test.	The development of practical video-based facial recognition applications requires additional work.
[2]	Daniel Schofield, Arsha Nagrani	A deep convolutional neural network (CNN) method is used and provided a completely automated pathway for the identification of faces in long-term video recordings of wild chimpanzees.	The total success rate for both name recognition and sex categorization was 92.5%. They created co-occurrence matrices using the recognised faces to track changes in the social network structure of an ageing population.	longitudinal video archives and their promise to address fundamental issues in conservation and behaviour.
[3]	Guodong Guo Na Zhang	The strategies for recognising faces using deep learning were given and reviewed in detail. About 330 papers in this field have been summarised that suggest various deep learning techniques. It examined key deep learning theories present a brief summary of studies on problems relevant to face image analysis and face recognition, such as adapting for changes in stance, age, lighting, mood, and heterogeneous face matching.	Deep approaches have shown to be effective in a variety of areas, including the processing of RGB-D and video in addition to heterogeneous face matching. We have also offered face databases that include related information, photos, videos, and to different types of face data, enabling cross-modal FR. Many significant steps forward have been made because of developments in face recognition technology, even if it can only work with static images.	In practise, there are still some challenges.

[4]	Yi-Chen Chen, Vishal M. Patel	The author developed the idea of face recognition using video dictionaries, which generalizes the work done with sparse representation as well as face recognition using still image dictionaries. Unrestricted video sequences are the goal of the Face as well as Ocular Challenge Series (FOCS) Video Challenge, was used by the author to illustrate their methodology.	According to experimental findings on three separate datasets, individuals perform best at face and body recognition while observing the subject in motion.	Next-generation video identification algorithms should make effective use of both facial and bodily features.
[5]	Hao yang and xiaofeng han	The author primarily establishes four directions to consider the issues: Face recognition attendance monitoring system interface settings; actual check-in accuracy rate of the a face recognition system; face recognition attendance system absenteeism rate. In conclusion, we suggest implementing a face recognition system.	In experiments, The video face recognition system's success rate has been calculated, and it stands at 82% at present. The facial recognition attendance system may minimize the time it takes to check in by roughly 60%. The phenomenon of pupils leaving class early and skipping it has significantly decreased in frequency.	The system has achieved significant advancements that have significantly increased attendance rates and the accuracy of facial recognition technology. It merits further scientific investigation and understanding.
[6]		For video face recognition, the author proposes a new approach When several face photos are fed into a component-wise feature aggregation network (C-FAN), and a single feature vector is generated to be utilised for recognition. To train the whole network, there are two stages: To recognise faces in videos, I first train a standard convolutional neural network (CNN) on still	The proposed C-FAN network effectively aggregates data to reach state-of-the-art performance feature vectors from across all video frames to generate The results of experiments on 3 benchmark datasets YouTube Faces, IJB-A, as well as IJB-S—show that the 512-dimensional feature representation is compact enough to represent merely a	Future research will look at an aggregation network that combines several degrees of fusion.

		images, and Then, In order to gather the quality score for every feature piece, I have an aggregation module added to a original network.	video sequence.	

III. METHODS

A. Eigen-faces:

A collection of eigenvectors called Eigen-faces is used to the computer vision issue of recognizing human faces. They make reference to an appearance-based method of face recognition that aims to identify variations among a group of face photographs and then uses this data to encode and contrast pictures of different faces.

The issue of precisely which components of the face stimulus are crucial for identification has been largely disregarded in earlier automated face recognition research, which assumed that predetermined measures were adequate and relevant. This gave rise to the idea that coding and decoding face photos according to information theory would help us understand the information contained in them by highlighting important local and global "features." It's possible that these traits have little to do with how our eyes, nose, lips, as well as hair are seen. The objective is to use information theory to facial images in order to extract the most relevant information, encode it, and then compare it to a database of encoded models. Unaffected by any evaluation of features, capturing variance in a group of face photographs and The information contained in a face picture may be easily extracted by utilising it to encode as well as compare several versions of the same face. Either the main components of a distribution of a collection of face photographs or the eigenvectors of their covariance matrix must be found. These eigenvectors may be thought of as a set of characteristics that, when combined, characterise the diversity in face images. Since each pixel in the picture contributes differently to the total eigenvector, we may think of the Eigen face as a form of spectral face.

It's possible that a combination of Eigen faces may represent each face picture in the training set. There are as many possible Eigen faces as there are images of human faces in the training set. You may get a close approximation of the faces by using just the "best" Eigen faces, which ones have the largest eigenvalues and hence explain the most variation in the collection of faces. The efficiency of computing is the main justification for employing fewer Eigen faces. The largest M' Eigen faces occupy a region of "face space" that is M' dimensions in size. Eigen functions of linear systems as well as the fundamental functions of the a Fourier decomposition (sinusoids with varying frequency and phase) form the basis vectors of a Eigen face decomposition.

The stages involved in recognition are outlined below:

- Initialization: The Eigen faces, which characterise the face space, cannot be calculated without first obtaining the training dataset of facial images.
- When a new face picture is encountered, the input image is projected onto each of the M Eigen faces to generate a set of weights.

- Check to see whether the picture is sufficiently near to "face space" to determine if the image is indeed a face (whether known or unknown).
- When a face is detected, the individual's unique weight distribution is classified as either known or unknown.

B. Neural Networks:

- **Feature extraction:**

Face representational data or feature vectors are the target of feature extraction. There are three primary techniques for obtaining features: Three common statistical methods are principal component analysis (PCA), linear discriminant analysis (LDA), as well as Fisher discriminant analysis (FDA).

- **Principal Component Analysis (PCA):**

The information theory method is used to guide the usage of PCA for face recognition. It effectively extracts the pertinent data from a face image and encodes it. The training face image data's subspace in the image space is identified, and the pixel values are de-correlated. The traditional depiction of a face is achieved by projecting the original face picture onto a coordinate system established by its basic components. Information compression, decorrelation, and dimensionality reduction are all achieved by projecting facial images into the main component subspace, which aids in decision-making. For the most relevant components in the face distribution, one may use mathematics to search for eigenvectors of the covariance matrix of a collection of face images.

- **Linear discriminant analysis (LDA):**

In machine learning, linear discriminant analysis is used to find the best linear features to use when dividing data into several groups. Next, a linear classifier is built from the resulting combinations. Since a multidimensional area is mapped into the a space with fewer dimensions, this is also considered feature reduction before further categorization. Linear discriminant analysis is often used in contexts involving classification. In the case of face recognition, for example, the numerous pixels that make up a face are first converted into a smaller number of linear combinations so that they may be sorted into appropriate categories. Fisher-faces are the linear combinations that can be found using LDA. Face identification uses linear discriminant analysis (LDA), whereas eigenfeature discriminant analysis is used for face retrieval. The LDA is a method for representing an image that draws attention to its discriminatory character by projecting it onto a system of fisher-faces with nonzero eigenvalues. In LDA, we choose the linear subspace which maximises the quotient:

$$\frac{\left|\Phi^T S_b \Phi\right|}{\left|\Phi^T S_W \Phi\right|}$$

$$S_b = \frac{1}{c}\sum_{k=1}^{c}\left(\mu_k - \mu\right)\left(\mu_k - \mu\right)^T$$

is the between-class scatter matrix, as well as

$$S_w = \frac{1}{M}\sum_{k=1}^{c}\sum_{i|x_i \in C_k}\left(x_i - \mu_k\right)\left(x_i - \mu_k\right)^T$$

the within-class scatter matrix,

where c is the client count, M is the sample size of face photos used for training, x_i , μ is the grand mean, as well as μ_k is the mean of class C_k .

- **Fisher Discriminant Analysis (FDA):**

One of the most important feature extraction methods is Fisher discriminant analysis. Fisher discriminant analysis is often used in two different ways: the Foley-Sammon linear discriminant analysis (FSLDA) as well as the uncorrelated linear discriminant analysis (UCLDA). Studies demonstrate that each FSLDA discriminant vector's Fisher criteria value is consistently greater than the value of the equivalent ULDA discriminant vector. This might be seen as a benefit for FSLDA as a higher Fisher criteria value for a discriminant vector indicates stronger discriminability. While FSLDA often gets correlative feature components and perhaps extremely correlative feature components, ULDA invariably receives the uncorrelated feature components. It seems that discriminant vectors perform better when the feature components they extract from data are less correlated to one another. Therefore, removing uncorrelated feature components provides a benefit for ULDA in this regard. The best discriminant vectors seem to also correlate to the lowest correlations between the retrieved feature components and the maximum Fisher criteria values. Unfortunately, neither ULDA nor FSLDA can provide such perfect discriminant vectors. Additionally, neither the FSLDA nor the ULDA are particularly effective. We may assume that, for discriminant vectors, there is an equilibrium between values on the Fisher criteria that are high, low correlation between recovered features components, with excellent discriminant vector performance.

C. Neural Networks:

Neural networks are a very efficient and trustworthy classification approach that may be used to make predictions both for known and unknown data. It performs well on nonlinear and linear data sets that can be kept apart. Voice recognition, face recognition, fingerprint recognition, iris recognition, as well as scene interpretation are just some of the many uses that have been found using NN. Synthetic "neurons," or "nodes," form the backbone of an ANN. These nodes are connected by edges, and the strength of the excitation (maximum value +1.0) or inhibition (minor value -1.0) at each node is indicated by its connection (maximum value -1.0). (a -1.0 maximum is possible) High values for connections stand for robust ties. Each node is built with a transfer function built in. An ANN consists of three distinct sorts of neutrons: input nodes, hidden nodes, as well as output nodes.

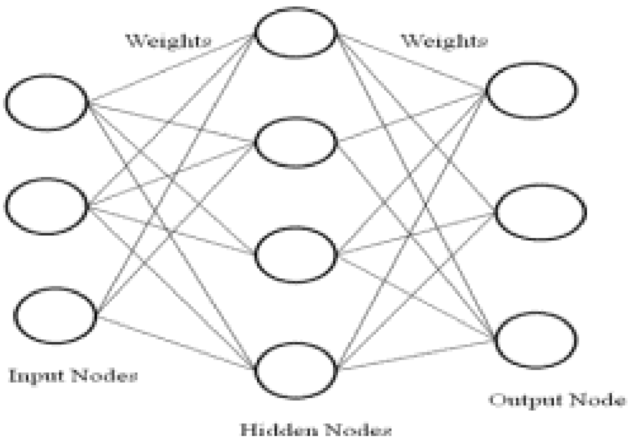

Figure 1NN

Data may be entered as numerical expressions into the input nodes. Activation values are used to display the information, with larger values indicating more robust activity at the corresponding nodes. The data then arrives to the network. The amount of activation that is transmitted from node to node depends on a number of factors, including weights, inhibition or excitation, and transfer functions. Values of activation received from each node are added, and then the sum is scaled up or down depending on the node's transfer function. Before reaching the final output nodes, the activation passes via the network's "hidden layers." As a result, the meaningful is mirrored in the nodes of the output graph.

D. Artificial Neural Network:

Gender categorization, face recognition, and expression classification are just some of the pattern classification problems that ANN machine learning has been used to. An ANN classifier offers excellent generalization and strong learning capabilities, which are benefits for classification. The need to oversimplify the classifier is avoided when an ANN is used, because it may be taught, using the features vector as input, a sophisticated mapping for classification. Face recognition systems that make use of neural networks and other learning approaches have been implemented owing to their promise for greater generalisation.

The artificial neural network (ANN) paradigm used here is the multilayer feed-forward network (MFNN). For those unfamiliar, an MFNN is a subset of non-linear networks that consists of inputs (the input layer), hidden non-linear neurons, as well as an additional non-linear output layer. MFNN's highly adaptable non-linear structure makes it a powerful tool for dealing with a broad range of difficult pattern recognition as well as regression issues. As a supervised learning technique, Showing the network examples from the a training set over and over again while modifying the neural weights to obtain the desired output is what error-correcting back-propagation is all about. How you train a network to perform a specific task. For the sake of the greatest possible descent, the weight matrices in a gradient descent algorithm are adjusted. It's important to choose η, the learning constant, carefully. The method may frequently overshoot the answer if it is too big, meaning that convergence may be slow or may not occur at all. If it's too little, the algorithm will move very slowly toward the answer, which will again cause a delayed convergence and raise the likelihood that the algorithm would become trapped in local minima. Momentum and adaptive learning are the two basic strategies for solving these issues. If we are continually moving in the same way as the momentum technique dictates, then we want to develop some momentum in that direction. If successful, this will help overcome any small local minima as well as speed up the convergence process.

Standard: $\Delta w(t) = -\eta \nabla E(t)$

Momentum: $\Delta w(t) = -\eta \nabla E(t) + \alpha \Delta w(t-1)$

where α is the momentum term.

To prevent overshooting, It is common practise for adaptive learning rates to begin with a high value and gradually decrease as we get closer to the answer. The feature extraction module's output data serves as the training's input data.

Euclidean distance (E.D.):

The closest mean classifier, often known as the Euclidean distance, is a common measure of distancing in the context of decision rules.

$$d_E(x, w_k) = \sqrt{(x - w_k)^T (x - w_k)}$$

Fuzzy classification:

The fuzzy set theory-based computer framework known as fuzzy inference systems (FIS) has been effectively used in a variety of applications. Success is primarily attributable to their compatibility with human perception and reasoning, as well as to the fact that they are easy to use and intuitive, all of which are crucial for system acceptability and usage. Unit fuzzification, rule base, inference engine, and defuzzification are crucial components of a fuzzy inference system. Fuzzification is the process of changing a variable's clear value to a cryptic one. Several fuzzy rules make up the rule basis. The inference engine calculates the outcome of a blanket application of all rules using the current values of any and all fuzzy variables. The defuzzification unit determines the defuzzified value of the inference engine's total fuzzy output.

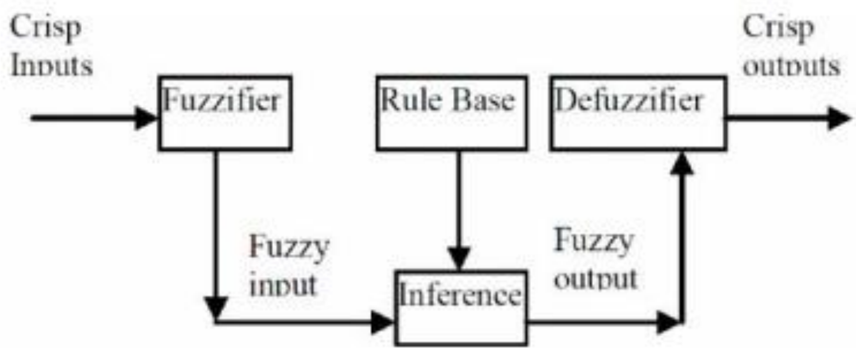

Figure 2 FIS

Recognition system:

Several 2D images will be gathered while video is being acquired using a camera. One or more persons may be seen in every image. The initial phase is segmenting the images in order to identify areas of interest that correlate to the various faces in each frame before doing face detection. Once an area of interest has been identified, Next, we proceed to the classification phase, which enables us to assign a percentage of likelihood that each recognised face belongs to each face class. The figure illustrates the fuzzy recognition system's organisational structure.

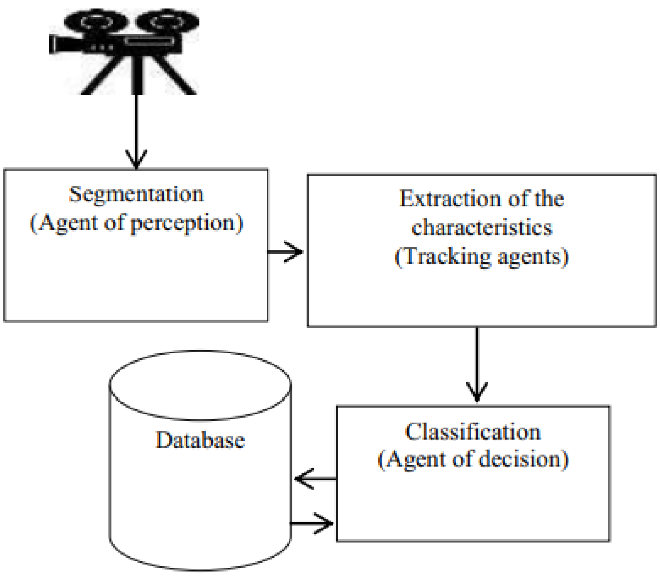

Figure 3 Structure of Fuzzy recognition system

Structure of the fuzzy controller:

Following is a formulation of the fuzzy problem of face recognition:

Let $C = \{C_1; C_2; \ldots; C_m\}$ be a collection of m face classes that are kept in a database, with m standing for the total number of classes.

Let X_j be a face j's characteristic vector containing d real values.

$X_j = (e_1; e_2; \ldots e_d) \in \Re^d$

$f : \Re^d \rightarrow [0,1]^m$

$f(x_j) = (\mu_1(X_j); \ldots; \mu_m(X_j))$

Where $\mu_i(X_j)$ represents the degree of membership of the face X_j to the class C_i

A system called a fuzzy controller has many inputs and many outputs. Each set of input values has a unique set of output values in this controller. Giving each identified face j a vector D_j with real values between 0 as well as 1 describes the degree to which the face j belongs to each class C_i in this fuzzy clustering approach.

The figure below illustrates the fuzzy controller's structure.

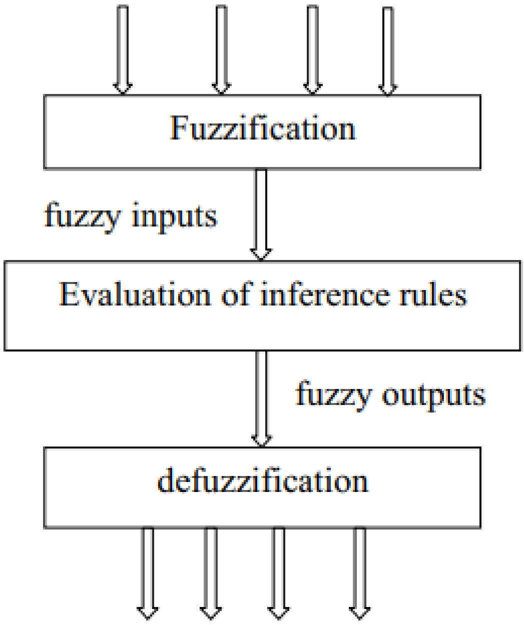

Figure 4 Structure of fuzzy controller

Three matrices M_k $(k = 1-3)$, whose column shows faces and those whose lines represent face classes, are produced via this fuzzy classification.

$$M_k = \begin{bmatrix} D_{c_1}(x_1) & \cdots & D_{c_1}(x_n) \\ \cdots & \cdots & \cdots \\ D_{c_m}(x_1) & & D_{c_m}(x_n) \end{bmatrix}$$

Where,

M_1 : Matrix of the fuzzy set "bad similarity"

M_2 : Matrix of the fuzzy set "average similarity"

M_3 : Matrix of the fuzzy set "good similarity"

$D_{c_i}\left(x_j\right)$: Degree of membership of the face x_j to class C_i

With: $$0 \le D_{c_i}\left(x_j\right) \le 1 \; and \; \sum_{i=0}^{m} D_{c_i}\left(x_j\right) = 1$$

$$D_{c_i}\left(x_j\right) = \frac{dist\left(x_j, c_i\right)}{\sum_{k=1}^{m} dist\left(x_j, c_k\right)} \quad pour 0 \le j \le n-1$$

Where:

m: represent the number of classes,

n: represent the number of face t classify,

c_i: is the characteristic vector of the center of gravity of the class i,

dist (x_j, c_i): represent the Euclidian distance enter a face x_j as well as the center of gravity of the class c_i.

Fuzzification involves applying the preceding formula on the matrix M_k in order to calculate it. It permits the conversion of digital data into language variable.

Each face in our article is defined by four factors.

- The skin tone
- The horizontal eye-width (Hb). The (D1) Distance
- The distance between the lips as well as the eyes (D5 et D4)

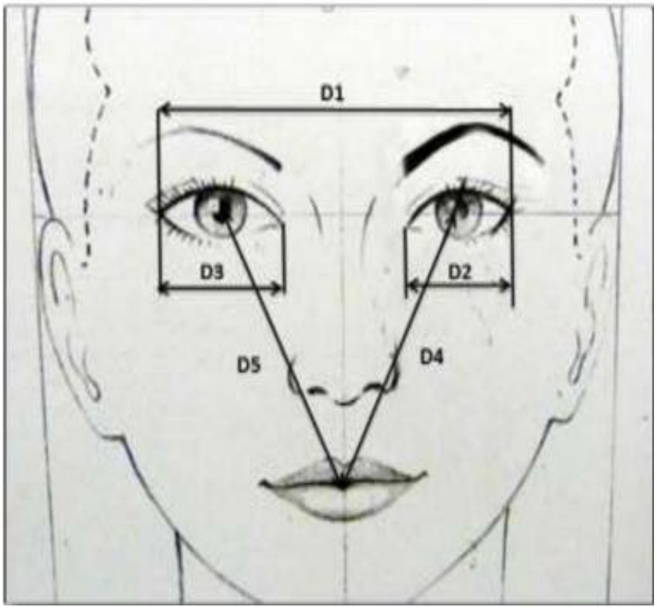

Figure 5 The parameters characteristic of the face

Since there might be several fuzzy subsets of an input value, the purpose of the fuzzy inference process is to calculate the relative degrees of membership for each. A facial membership function for each fuzzy collection is therefore unnecessary. Good similarity, medium similarity, as well as awful similarity fuzzy sets were used.

Genetic algorithm:

It is a kind of a random search algorithm motivated by both genetics and natural selection. After each cycle, a pool of potential answers is maintained by simulating biological reproduction, crossover, and mutation, and On the basis of some measure, the best people are selected from the groupings of solutions. Genetic operators combine individuals to produce fresh generations of solutions until the convergence index is reached.

A genetic algorithm consists of four main components: the genetic operators (selection, crossover, as well as variation), the fitness function, the coding method, and the settings for control. The potential answers to the issue will be encoded into chromosomes, or persons, when a genetic algorithm is employed to solve it. A group of people get together to produce the first solution. The people meeting the termination requirements may be output after computing the fitness function, and the procedure is then complete. If not, people will mix, mutate, and recombine to create the subsequent population. The previous generation's positive traits are passed down to the next population, which is superior to it, allowing it to steadily progress toward a better solution.

It is clear from the previous introduction that genetic algorithms provide a generic framework for resolving complex system optimization issues that is independent of the area and kind of issues. Fitness function is the foundation of the genetic algorithm. Rearranging the population's unique structure may be done in cycles by applying genetic operations to its members. There are many straightforward genetic algorithm models that fit these descriptions:

As a formula: C- unique coding approach

E-Individual fitness evaluation functionality

P0- initial population

N- Population size

U- Selection operator

C-Crossover operator

W-Mutation operator

T- Genetic operation's termination condition

1. Initial population coding and generation

Select the best method of coding for the problem at hand. As well as create a population of N chromosomes with a predetermined length at random.

$$f_i = fitness\left(pop_i(t)\right) \quad pop_i(t), t=1, i=1,2,.....N$$

2. Evaluation of Health and Fitness Potential

Each chromosome's fitness in population pop (t) is determined by its fitness value, tpop (t).

$$f_i = fitness\left(pop_i(t)\right)$$

3. Assess whether or not the convergence condition holds. The procedure will continue if the search results returned are appropriate.

4. Select operation

Each person's likelihood of being chosen is based on their level of fitness.

$$p_i = \frac{f_i}{\sum_{i=1}^{N} f_i}, i = 1,2,3,....N$$

Next-generation populations are generated by randomly selecting individuals from the above probability distribution, which is based on the current generation's population TPOP (t).

$newpop(t+1) = \{ pop_j(t) \mid j = 1, 2, ...N \}$ $newpop(t+1) = \{ pop_j(t) \mid j = 1, 2, ...N \}$

5. Cross operation

By mating with probability Pc, N chromosomes tcrosspop (t + 1) were produced.

6. Mutation operation

To cause chromosomal genes to mutate, use a lower probability Pm. The t mutpop (t + 1) population is created. The offspring of a geno-surgery are referred to as the pop (t) = mutpop (t + 1) population. It's the progenitor of the next genetic operation and loops back to 2 at the same time.

IV. REFERENCES

[1]. Z. Huang et al., “A Benchmark and Comparative Study of Video-Based Face Recognition on COX Face Database,” IEEE Trans. Image Process., vol. 24, no. 12, pp. 5967–5981, 2015, doi: 10.1109/TIP.2015.2493448.

[2]. D. Schofield et al., “Chimpanzee face recognition from videos in the wild using deep learning,” Sci. Adv., vol. 5, no. 9, pp. 1–10, 2019, doi: 10.1126/sciadv.aaw0736.

[3]. S. A. Nazeer, N. Omar, and M. Khalid, “Face recognition system using artificial neural networks approach,” Proc. ICSCN 2007 Int. Conf. Signal Process. Commun. Netw., no. June, pp. 420–425, 2007, doi: 10.1109/ICSCN.2007.350774.

[4]. Y. C. Chen, V. M. Patel, P. J. Phillips, and R. Chellappa, “Dictionary-based face recognition from video,” Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 7577 LNCS, no. PART 6, pp. 766–779, 2012, doi: 10.1007/978-3-642-33783-3_55.

[5]. H. Yang and X. Han, “Face recognition attendance system based on real-time video processing,” IEEE Access, vol. 8, pp. 159143–159150, 2020, doi: 10.1109/ACCESS.2020.3007205.

[6]. S. Gong, Y. Shi, N. D. Kalka, and A. K. Jain, “Video Face Recognition: Component-wise Feature Aggregation Network (C-FAN),” 2019 Int. Conf. Biometrics, ICB 2019, 2019, doi: 10.1109/ICB45273.2019.

[7]. (2002) The IEEE website. [Online]. Available: http://www.ieee.org/

[8]. “PDCA12-70 data sheet,” Opto Speed SA, Mezzovico, Switzerland.

[9]. A. Karnik, “Performance of TCP congestion control with rate feedback: TCP/ABR and rate adaptive TCP/IP,” M. Eng. thesis, Indian Institute of Science, Bangalore, India, Jan. 1999.

[10]. J. Padhye, V. Firoiu, and D. Towsley, “A stochastic model of TCP Reno congestion avoidance and control,” Univ. of Massachusetts, Amherst, MA, CMPSCI Tech. Rep. 99-02, 1999.

[11]. Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specification, IEEE Std. 802.11, 1997.