International Journal of Advanced Research in Engineering and Technology (IJARET)
Volume 11, Issue 12, December 2020, pp.1813-1829, Article ID: IJARET_11_12_171
Available online at http://www.iaeme.com/IJARET/issues.asp?JType=IJARET&VType=11&IType=12
ISSN Print: 0976-6480 and ISSN Online: 0976-6499
DOI: 10.34218/IJARET.11.12.2020.171

FEATURE RECOGNITION: A CONTEMPORARY SURVEY OF GROWING NEEDS BLENDED WITH MACHINE LEARNING FOR SURVEILLANCE, SECURITY AND INTELLIGENCE SYSTEMS

M. Swetha
Computer Science of Engineering, Keshav Memorial Institute of Technology,
Narayanaguda, Hyderabad, India

ABSTRACT

Feature recognition is one of the most prominent areas of Machine Learning (ML) since ages together human are fascinated with the world of colors, features, virtues, and such artifacts for gaining inside into the knowledge sharing. The intuitiveness of the human mind is quite capable of recognizing and pursing miniature details about the surroundings and the environment where they live in. In so far, even the animals and other living beings also have the power to recognize and assimilate the information—however, they lack in understanding, interpolation and other feature extraction details.

The traditional AI has been used for various applications starting from feature detection, extraction and applying convolution techniques for feature engineering. In achieving so Principal Component Analysis (PCA) gives a mathematical model to get inside with reduction feature extraction and computational efficiency. This image processing area has being fascination for humans to see nice features and increase their happiness quotient.

On the other side, the people are playing the game to reshape the feature and get a noticed like criminals, terrorists, fraudulent and other such irrelevant human behavior. However some of the features never changes through the life span of like retina, lines or regions on the hand, thumb impression etc., This salient features are going to be retain throughout the life journey and everlasting.

In this article an attempt is being made to use ML algorithms for such areas and with pertinent mathematical background; thereby we can understand certain artifacts of features and use them for a good cause.

Key words: Feature Detection, Extraction, PCA, Eigenvectors, Eigen Features.

Cite this Article: M. Swetha, Feature Recognition: A Contemporary Survey of Growing needs blended with Machine Learning for Surveillance, Security and Intelligence Systems, *International Journal of Advanced Research in Engineering and Technology*, 11(12), 2020, pp. 1813-1829.
http://www.iaeme.com/IJARET/issues.asp?JType=IJARET&VType=11&IType=12

1. INTRODUCTION

Face recognition is a vast field comprising of enormous processing involving high degree of matrix operations with the advancement in technology. Every passing year offers solutions with varying and increasing complexity. Computer based face identification is the hall-mark application of generic image processing. With the use of modern techniques and algorithm, varies operations like reconstruction, enhancement, compression, rendering, segmentation, feature extraction, face detection and face recognition can be accomplished in a better way.

Looking back in years, the Digital Image Processing (DIP) has gained immense populating among researches in last 3-4 decades. Further, the face recognition system is identifying the face, is closer to humans imaginative comprehension. Towards this, many digital systems have been evolved over these years and numerous methods and techniques have been proposed in the DIP literature. Meaningfully, the face recognition process comprises of person's face and relative these extract features to match with the database. This capability enhances the usefulness of the technology[1].

Furthermore, as machine learning algorithms are able to identify the pattern underlying the given dataset, recognition of pattern on human face is well addressed domain of ML. This system has taken the equivalent role of human is identifying and comprehensing the features changes due to various facts, like aging, environment, and alike. However, the basic application domain typically includes:

- Identifying a driver for a valid driving license
- Portrait certification
- Signature authentication for bankcheque veracity
- Safety measures
- Image dispersion, and many more.
- Identifying a suspect in the crowd and if found take security measures.

The elusive 2 significant artifacts while doing a pattern recognition examination of facial features. The key idea pertains to use of neural network (NN) incorporating the recognition of face [2]. Elastic graph match method for face recognition [3], Eigen face [4]. Additionally the feature extracts from the face, focuses on facial extraction, and template matching approach [5, 6].

Ironically, the face recognition techniques involve 3 phases: namely (i) face recognition, (ii) features extraction and (iii) face detection [7]. In this study a survey of these 3 techniques are presented with an emphasis on face detection from the front view face images founded on skin color. The process is exciting one having an order form as extracting face region detection and lastly eyes extraction, the features like lips and nose are extracted from the distance between the two eyes. Subsequently following this phase, a small face region is selected in order to increase the execution period. The left eye is extracted from the features region. For this cognition purpose, blended with *Eigen features* and *Fisher features* are used. There are numerous varied methods proposed in the viewpoint of face recognition in the literature. A great amount of work has already been done under this fascinating area for several decades and various methods have been identified and the purpose of this article is to highlight about those prevailing techniques, approaches [4], Gabor wavelets [8] to 2D distinct cosine transform for facial features extraction [9], geometrical approach [10] PCA based facial features extraction [11] has a disadvantage that the method is delicate to the illumination changes.

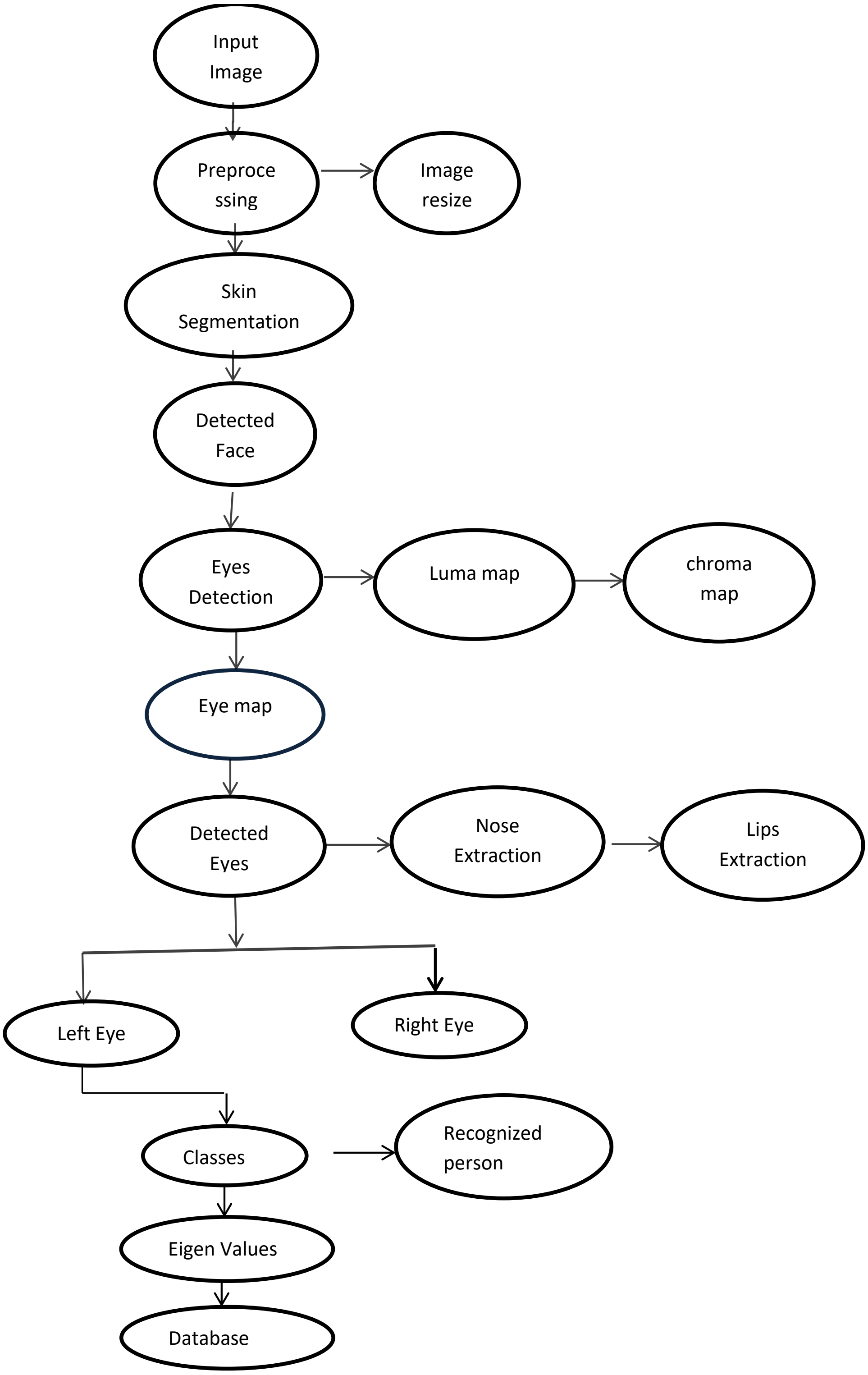


Figure 1 The process flow of feature detection and extraction

The most popular and cited methods in the DIP literature are—Principal Component Analysis (PCA) [4], Independent Component Analysis (ICA), Linear Discriminator Analysis (LDA) [11], Neural Networks, Kernel methods, and Gabor Wavelets [12]. Principally there are two research fronts; the first one with progress in facial features recognition been taken as an exciting avenue and the other one has taken Eigen and Fisher values for features extractions. Nevertheless, the efforts put by both group views are equally taken as active part in the progression in facial detection systems in general. The latter groups mainly concentrate on human features, such as eyes, lips, and nose in order to recognize a facial image. In doing so, an accurate facial feature s extraction is very important and it definitely addresses different issues.

Foregoing, the approach proposed by Yuille et al. (1989) uses face features extraction as of vital values and chiefly used templates of deformable parameters to extract mouth and eyes. However, this approach has been computationally expensive and gave less accurate results. Later on, Wiskott et al., (1999) proposed another method that extracts facial features and performs recognition based on template matching approach. For extraction, the algorithm first locates the facial area looking at skin color subsequent to the extraction of applicant's iris centered calculated cost and later the mouth region is extracted by use of RGB palette. Further, the associated constituent process is utilized to precisely extract out the lips region. To ensure that the detected lips are mouth, the mouth's curve points are extracted using SUSAN approach. Finally the face identification process is performed using template matching process.

2. BACKGROUND AND RELATED WORK

2.1 Image Processing

In this section, the background with pertinent details is presented in order to augment the process and highlighting some conceptual mathematical derivations needed for exploration.

Images are the evident sources in image processing applications. Image processing will dramatically change the human computer interaction in future. A large number of image processing applications, tools and techniques help to extract complex features of an image. With the advancement in CPU processing capabilities and memory technology, the present image processing works very well with multi-dimensional features as well as in real time streaming.

An image is defined by a function called " *Two-Dimensional image space* " and can be specified by spatial coordinates as plane coordinates and its pair coordination are amplitude, which may represented as ***f* (x,y)** , where the intensity of these coordinates are gray level images as shown in Fig. 2.

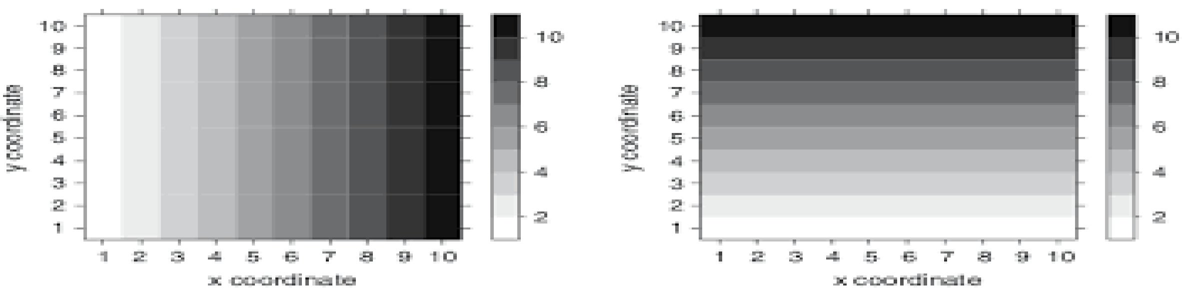


Figure 2 An Gray scale image with scale of variations in intensity

As we can see from Fig. 2, when the spatial coordinates and the intensity values are finite with discrete quantities, then it is called as Digital Image Processing (DIP). The digitized images are easier to analyze and manipulate in order to improve the quality with the use of some mathematical calculations. Here the image is given as input, parameters of an image as processed through algorithms and characteristics related to the images as an output. Digital Image Processing (DIP) is the process of digital images through various algorithms. This digital image processing has been employed in number of areas such as pattern recognition, remote sensing, image-sharpening, color and video processing and medical imaging. An elaborated discussion is given by Ranu Gorai (2016) in [13] about how an image and the picture are defined in our daily life and lucidly explained 3 methods , namely (i) sharpening the edges,(ii) noise removal and (iii) removing the motion blur for an image. Further , Jangala Sasi Kiran (2015) et al., [14] have proposed an algorithm to improve the overall accuracy of the hand written character under pattern recognition field using image processing techniques like feature extraction, image restoration and image enhancement . B.Thamotharan et al., [15] proposed that the digital image processing concepts are done by different algorithms and highlighted noise and edge detection algorithm. They discussed other two concepts, viz., mean and median filtering for radiographic images and compared them [16], proposed that this paper mainly focusing on the security based system by modern digital image processing also this paper gives a way to process a videos from variety of video devices. First set the continuous frames from the videos then it had been processed under SUSAN for extracting the features [17] exposed about the data clustering for clustering of objects. A specific algorithm is used for group detection from an image using distance metrics through linear features [18] proposed that the main contribution of this research is plant phenotyping research or the automation of research in image acquisition to minimize the data at geo referencing errors and for modular data visualizations and also fastest data collections [19],discussed about accelerating performance on FFT based image processing algorithm between Central Processing Unit (CPU) and Graphics Processing Unit (GPU) [20] proposed an age estimation algorithm for facial images using multi-label sorting. They changed the tedious process in traditional multi valued classification algorithm and compared the results with some classic algorithms of age estimation by verified the efficiency and accuracy of the algorithm.

2.2 Mathematical Techniques

This section highlights the fundamental techniques used in the face identification and recognition. An overview will augment the understating of the theme.

Correlations:

Correlations are frequently measured in terms of correlation coefficients, i.e., a measure of the degree of association between two variables. The most commonly known is the Pearson coefficient ‘r’ —whose value indicates how strong a correlation is and can vary from −1.0 to + 1.00. An ‘r’ value of +1.0 indicates an ideal positive correlation, and −1.0 indicates just opposite. Nevertheless, in both these cases, the value of one variable can be predicted exactly for any value of the other variable. Importantly, intuitively, a 0.0 value signifies no relationship between the variables at all.

The computation of correlation coefficient involves going through a series of regular steps. These allow us to establish whether high scores on one variable are associated with high scores on the other, and if low scores on one variable are associated with low scores on the other. Eventually, as expected it turn out to be computationally intensive.

2.3 Eigenvalues and Eigenvectors

The Eigenface method is based on linearly projecting the image space to a low dimensional feature space [21, 22, 23].and uses principal components analysis (PCA) for dimensionality reduction yielding projection directions that maximizes the total scatter across all the classes, i.e., across all images of all the faces. While choosing the projection which maximizes total scatter, PCA retains redundant variations due to lighting and facial expression. Fig. 3 shows a PCA.

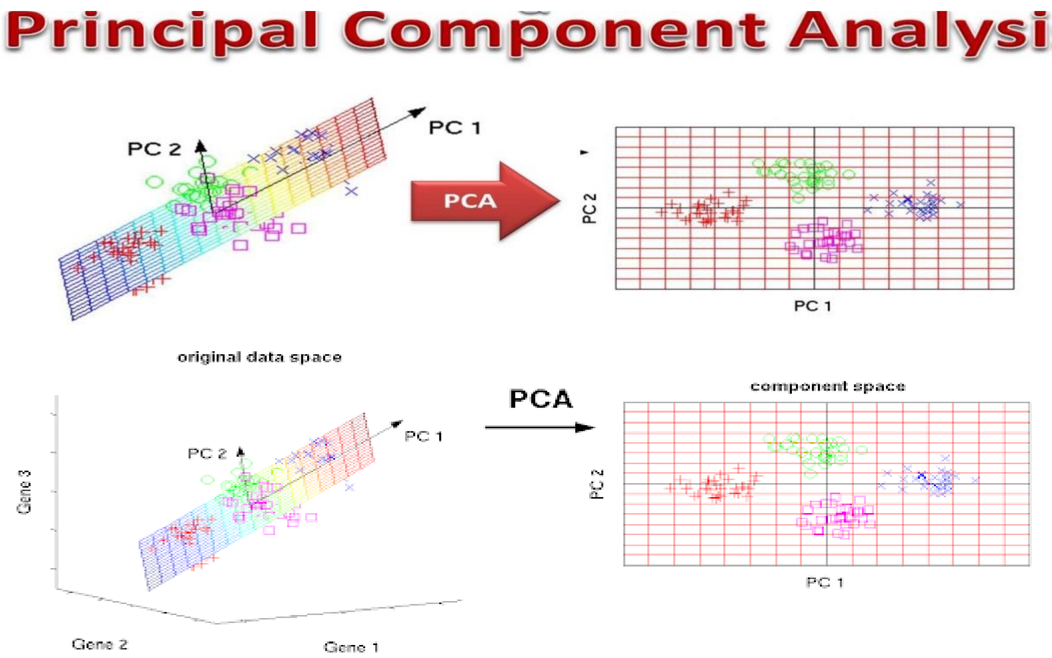


Figure 3 PCA

As noted earlier, the correlation methods are computationally expensive and require huge amounts of storage, it is obvious to go for a practice of dimensionality reduction schemes. The most commonly used for dimensionality reduction in computer vision— particularly in face recognition—is Principal Components Analysis (PCA) [21, 22, 23, 24,25], also known as Karhunen- Loeve methods, choose a dimensionality reducing linear projection that maximizes the scatter of all projected samples. More formally, considering a set of N sample images : { X_1, X_2, …. , X_N } taking values in n-dimensional image space, and assume that each image belongs to one of c classes $\{X_1, X_2, \dots, X_C\}$. Let us also consider a linear transformation mapping the original n-dimensional image space into an m-dimensional feature space, where m < n. The new feature vectors $y_K \in R^m$ are defined by the following linear transformation:

$$Y_k = W^T X_k \quad k = 1, 2, \dots, N \qquad (1)$$

Where $W \varepsilon R^{n\times m}$ is a matrix with ortho-normal columns.

If the total scatter matrix S_T is defined as

$$S_T = \sum_{K=1}^{N} (x_k - \mu)(x_k - \mu)^T$$

where n is the number of sample images, and $\mu \in R^n$ is the mean image of all samples, then after applying the linear transformation W^T, the scatter of the transformed feature vectors $\{y_1, y_2, \dots, y_N\}$

is $W^T S_T W$.

In PCA, the projection W_{opt} is chosen to maximize the determinant of the total scatter matrix of the projected samples, i.e.,

$$W_{opt} = \text{argmax} |W^T S_T W|$$
$$= [W1\ W2 \dots W_m] \qquad (2)$$

where $\{W_i | i = 1, 2, \ldots, m\}$ is the set of n-dimensional eigenvectors of S_T corresponding to the m largest eigenvalues. Since these eigenvectors have the same dimension as the original images, they are referred to as Eigenpictures in [21] and Eigenfaces in [22,23]. If classification is performed using a nearest neighbor classifier in the reduced feature space and m is chosen to be the number of images N in the training set, then the Eigenface method is equivalent to the correlation method.

2.4 Gabor Features and Extraction (1)

In image processing, a Gabor filter is a linear filter used for texture analysis, which essentially means that it analyzes whether there is any specific frequency content in the image in specific directions in a localized region around the point or region of analysis, as shown in figure 4. Frequency and orientation representations of Gabor filters are claimed to be similar to those of the human visual system (as per Wikipedia page). These representatives' characteristics have been found appropriate for texture representation and discrimination. In the spatial domain, a 2D Gabor filter is a Gaussian kernel function modulated by a sinusoidal plane wave.

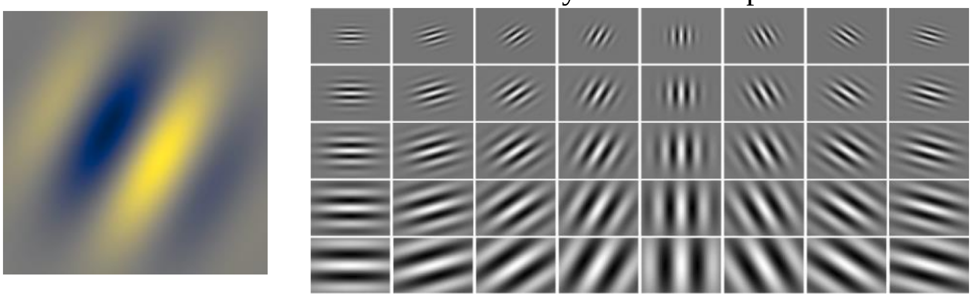


Figure 4 Gabor Features and Filter

So in the domain, we use these techniques — we first locate the feature points and then apply Gabor filters in each point in order to extract a set of Gabor wavelet coefficients. We use wavelet analysis because it can localize, in space-frequency, characteristics of images and it can represent faces in different spatial resolutions and orientations [26, 27]. For example, Zhang [28] has proposed a system of face recognition based on Gabor wavelet associative memory (a neural network)[29]. Wiskott [30] has proposed a face recognition system by elastic bunch graph matching representing the local features of faces (eyes, mouth, etc.) by a set of Gabor wavelets. The proposed system of face recognition can detect the face in an image and localize automatically 10 fiducial points. After that it characterizes the face by the geometric distances between the extracted points or by applying a set of Gabor wavelets (filters) in correspondence to each fiducial point. These two types of features extracted from face can be used either independently or jointly. They are used as input data to neural networks for classification. The recognition performances with different types of features are compared. The paper is organized as follows: the face detection system is described in section 2. Section 3 presents the algorithm of facial features localization and extraction of their characteristic points.

2.5 Feature Extraction and Feature Engineering

The next stage in order is features extraction from the received face as the input. Following this is to locate the features from detected face region. The extracted enumerated features are:

(i) Lips (ii) Nose (iii) Eyes

Figure 5 features showing human Nodal points

The eyes are located with chrome and luminance mapping and nose and lips are then extracted by means of a formula which is centered on the space existing among both eyes. We need seven following features , as shown in Figure 6, to be extracted from the faces which are:

Figure 6 Showing 7 features

- Lips center
- Lips left corner
- Lips right corner
- Nose tip
- Centre of the Left eye
- Centre of the Right eye
- Eyes midpoint

2.5.1 Eye extraction

To detect the eyes from a human face, the pertinent information of dark pixels on the face are required. Further, as we already know, the eyes color is usually different from the skin color, so using a color space and separating the colors would be a good idea in order to locate the eyes. The '*YCbCr*' color space provides modest information about the dark and light pixels present in an image. The color space called '*YCbCr*' has the tendency to establish the eye area with a high Cb and low Cr values. To detect the eye, another eye maps are constructed in the YCbCr color space which show the pixels with high Cb and low Cr values for eye region.

$$\text{Chroma Map} = 1/3\ (Cb/Cr + Cb^2 + (1-Cr)^2 \qquad (3$$

where Cb : chroma blue-difference

Cr : chroma red-difference

The luma map is built by applying the morphological (Sawangsri et al., 2004) operations of erosion-dilation on the luminance component of the YCbCr image. These operations are applied to enhance the brighter and the darker pixels around the eye region in the luminance constituent of the image.

$$\text{LumaMap} = \text{Ydialte} / (\text{Yerode} - 1) \qquad (4)$$

2.5.2 Nose and Lips detection

The nose and lips are extracted on the basis of distance between the two eyes–with the assumption that they lie at a specific ratio of the distance between the two eyes.

Nose detection

For nose extraction the distance between the two eyes is calculated as:

$$D = \sqrt{}\ (Lx - Rx)^2 + (Ly - Ry)^2 \qquad (5)$$

where

D = Distance between the center points of two eyes

L_x = Left eye x coordinate

R_x = Right eye x coordinate

L_y = Left eye y coordinate

R_y = Right eye y coordinate

After having the eyes distance a formula is generated on the basis of that distance to extract the nose. It is assumed that nose lies at a distance of 0.55 of the eyes distance. This is done as:

$$N = (My + D)*0.55 \quad (6)$$

where,

N = Nose tip point

M_y = Eyes midpoint

D = Distance between the center points of two eyes

Lips extraction

For lips extraction the three lips points extracted are:

- Lips center point
- Lips left corner point
- Lips right corner point

For lips extraction the distance between the two eyes is calculated as:

$$D = \sqrt{} (L x - Rx)^2 + (Ly - Ry)^2 \quad (7)$$

where,

D = Distance between the center points of two eyes

L_x = Left eye x coordinate

R_x = Right eye x coordinate

L_y = Left eye y coordinate

R_y = Right eye y coordinate

After having the eyes distance a formula is generated on the basis of that distance to extract the lips. For the lips center point, it is assumed that lips lie at a distance of 0.78 of the eyes distance. This is done as:

$$L = (M y + D)*0.78 \quad (8)$$

where,

L = Center point of lips

My = The midpoint of the space among both eyes

D = Space among the center points of two eyes

For the lips right and left corner point's extraction it is analyzed that the lips corners are almost located at a distance 0.78 of the eyes center points. These points are extracted as:

$$Ll = (Ly + D) * 0.78 \quad (9)$$

where,

Ll = Left corner point of lips

Ly = Left eye y coordinate

D = Distance between the center points of two eyes

Similarly the right corner point of mouth is extracted as:

$$Lr = (Ry + D) * 0.78 \quad (10)$$

where,

Lr = Right corner point of mouth

Ry = Right eye y coordinate

D = Distance between the center points of two eyes

3. BLENDING WITH MACHINE LEARNING

Whenever a human face has to be recognized, the feature extraction helps us intuitively. To enhance the image captured in the human mind, it might take spontaneous or deferred time to recall. Many a times has the age progresses, human are increasingly incapable of recollecting the features in a exact manner. With the progress in the machine learning algorithms, however it is a blessing in disguise such that a machine can remember (permanently or temporarily) an extracted feature. Since the inception of machine learning with deep neural network, it has indeed made the recognition and reconstruction fastly in order to have a very good memory recall. This feature is in fact a desirable proponent for a machine learning algorithm. Such a blending of ML with human cogniscience has prove to be very effective and the subsequent sections will highlight those blending's in simpler and institutive base.

3.1 Deep Neural network:

The face recognition process based on convolutional neural network (CNN) is primarily divided into 2 stages: namely (i) training stage and (ii) testing stage. In the first stage, a large number of face samples are collected to establish the training database, and the training samples are preprocessed.

Further, the preprocessing is usually divided into two parts in order to:

- Remove the noise of training data, normalization is needed,
- Make the distribution of training samples cover the illumination, scale and other situations as much as possible. We need to enhance the training samples to improve the generalization ability of the model.

A typical CNN layer is shown in Figure 5, where feature learning and classification are done layer wise.

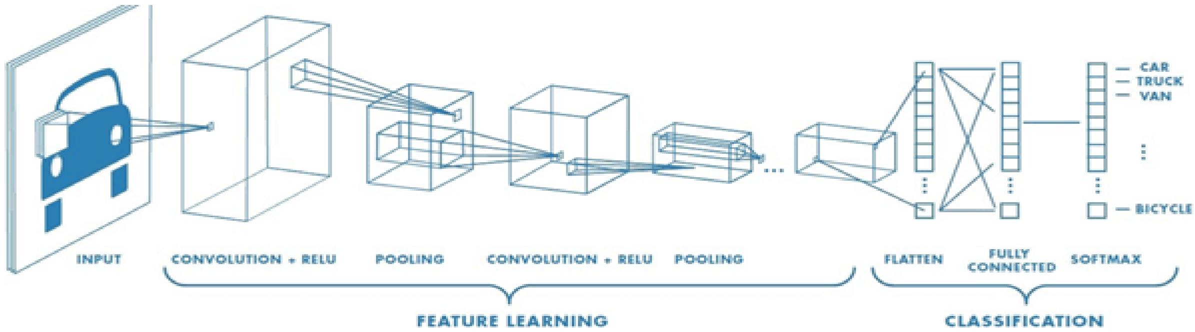


Figure 6 A CNN architecture depicting various layers

(Source: https://www.guru99.com/deep-learning-tutorial.html)

Subsequently, the training of multi-layer convolution network model is divided into forward propagation and backward propagation. The formers refers to that the training samples that are fed into the network model by batch for training, and the output eigenvectors are extracted through the feature extraction part. After the prediction through the classification layer, the predicted probabilities are found. Consequently, in the loss layer, the loss amount is calculated by using the loss function by comparing the predicted results with the given label errors.

The later one, i.e., backward propagation refers to the use of convergence function (such as gradient descent function) for the loss of forward propagation, which transfers the loss layer back to the feature extraction part, and updates the parameters of multilayer network layer by layer. The purpose of this function is to update the parameters of the network to their respective gradients, so as to reduce the loss gradually. That is to say, the training goal of the whole network model is to make the prediction results of the forward propagation of the model consistent with the given label results. After the training of network model, the feature vector of hidden layer has the feature of distinguishing face. Finally, face matching and recognition are carried out by measuring the feature vector. In the test phase, compared with the training phase, it is relatively simple. First of all, a face test database needs to be established. However, each sample in the test database usually appears in pairs. After the test sample is passed through the network feature extraction part, the feature vectors of a pair of faces in the test sample are obtained. Then, whether the pair of faces is the same person is determined by measurement and threshold.

CNN is mostly used when there is an unstructured data set (e.g., images) and the practitioners need to extract information from it for instance, if the task is to predict an image caption:

The CNN receives an image of let's say a cat, this image, in computer term, is a collection of the pixel. Generally, one layer for the grey scale picture and three layers for a color picture

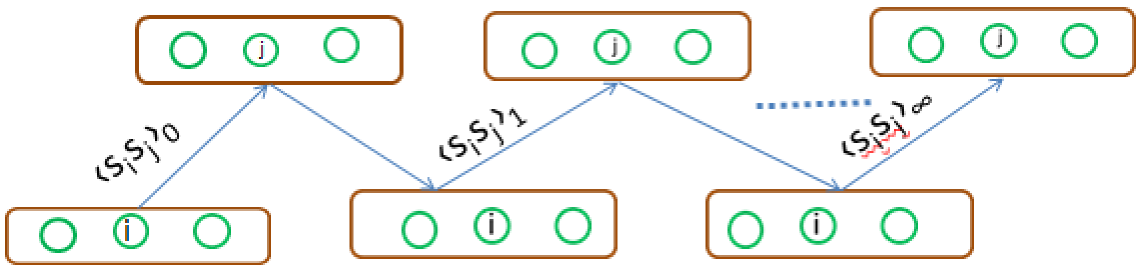
Figure 7 An image with different pixel colors and hue

During the feature learning (i.e., hidden layers), the network will identify unique features, for instance, the tail of the cat, the ear, etc.

When the network thoroughly learned how to recognize a picture, it can provide a probability for each image it knows. The label with the highest probability will become the prediction of the network. For more information on CNN, one can refer to excellent references in [36,37].

3.2 Boltzmann Machine (2)

Conventionally, a Boltzmann machine is a type of binary pairwise Markov random field (undirected probabilistic graphical model) with multiple layers of hidden random variables. It is a network of symmetrically coupled stochastic binary units. It comprises a set of visible units and layers of hidden units.



data time= 0 | reconstruction 1 | fantasy ∞

Figure 8 A traditional Boltzmann Machine

A Restricted Boltzmann machine (RBM) [31] incorporates with a layer of visible units and a single layer of hidden units with no hidden-to-hidden nor visible-to-visible connections. Because there is no explaining away [32], inference in an RBM is much easier than in a general Boltzmann machine or in a causal belief network with one hidden layer. There is no need to perform any iteration to determine the activities of the hidden units, as the hidden states, S_j, are conditionally independent given the visible states, S_i . The distribution of s_j is given by the standard logistic function:

$$p(s_j = 1 | s_i) = 1 \div 1+\exp(- \sum_I w_{Ij} s_i) \qquad (11)$$

Conversely, the hidden states of an RBM are marginally dependent so it is easy for an RBM to learn population codes in which units may be highly correlated. It is hard to do this in causal belief networks with one hidden layer because the generative model of a causal belief net assumes marginal independence. An RBM can be trained using the standard Boltzmann machine learning algorithm which follows a noisy but unbiased estimate of the gradient of the log likelihood of the data. One way to implement this algorithm is to start the network with a data vector on the visible units and then to alternate between updating all of the hidden units in parallel and updating all of the visible units in parallel with Gibbs sampling. Figure 1 illustrates this process. If this alternating Gibbs sampling is run to equilibrium, there is a very simple way to update the weights so as to minimize the Kullback-Leibler divergence, $Q^0||Q^\infty$, between the data distribution, Q^0, and the equilibrium distribution of fantasies over the visible units, Q^∞, produced by the RBM [4]:

$$\Delta w_{ij} \,\alpha < s_i s_j > Q^0 - < s_i s_j > Q \qquad (12)$$

Where $<S_i S_j>Q^0$ is the expected value of $S_i S_j$ when data is clamped on the visible units and the hidden states are sampled from their conditional distribution given the data, and Q^∞ is the expected value of $<S_i S_j>$after prolonged Gibbs sampling. This learning rule does not work well because it can take a long time to approach equilibrium and the sampling noise in the estimate of $<S_i S_j> Q^\infty$ can swamp the gradient. Hinton [1] shows that it is far more effective to minimize the difference between $Q^0 \| Q^\infty$ and $Q^1 \| Q^\infty$ where Q^1 is the distribution of the one-step reconstructions of the data that are produced by first picking binary hidden states from their conditional distribution given the data and then picking binary visible states from their conditional distribution given the hidden states. The exact gradient of this "contrastive divergence" is complicated because the distribution Q1 depends on the weights, but this dependence can safely be ignored to yield a simple and effective learning rule for following the approximate gradient of the contrastive divergence:

$$\Delta w_{ij}\, \alpha < s_i s_j >Q^0 - < s_i s_j >Q^1 \qquad (13)$$

4. SURVEILLANCE, SECURITY AND OTHER ISSUES

There are 3 main characteristics of video monitoring system, vize: (i) massiveness, (ii) unstructured and (iii) low value density. The traditional analysis methods are real-time monitoring and video query and playback manually. However, with the explosive advances in video monitoring system, there are certain issues which need to be addressed:

- Data management mode,
- Data analysis application and other technologies

These issues are very critical from security viewpoint.

In order to accomplish these, we need to augment number of persons with limited display equipment, i.e., the number of front-end cameras and number of display screens and some other factors [33, 34]. Noteworthy, we need to assure that all those scenes were monitored accurately and efficiently in past 24 hours for surveillance purpose. In not doing so it is also difficult to find and trace the video surveillance image afterwards, and the human and material resources investment well exceed. For instance, as cited in [35,36] the process of detecting a major case to find a criminal suspect in the video surveillance, a local public security forces had used roughly 2000 police personnel to conduct video playback and manual search for more than ten hours every day. As reported, the total video viewing volume was equivalent to almost 830000 films, which consumed a lot of manpower and material resources. This indeed is a herculean task as well as leading to manual fatigue.

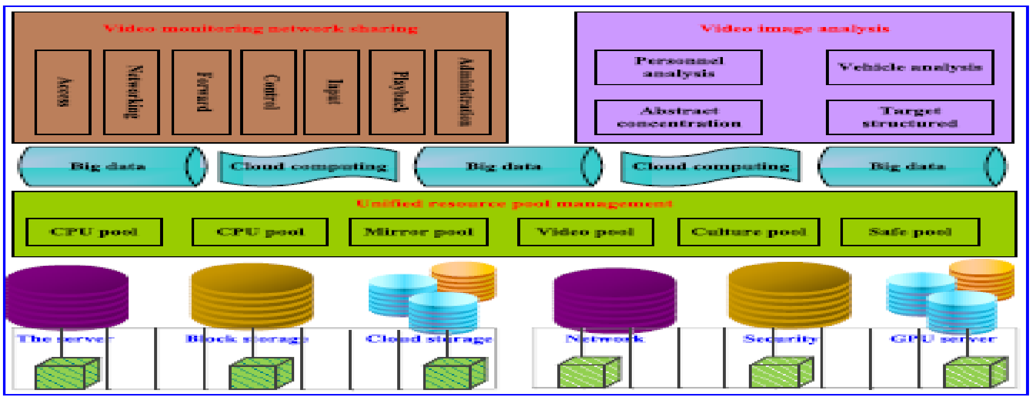


Figure 9 A hardware systems showing various components

Thus, to proliferate and offer ease of usage, in such scenario, the state-of-art IT expertise enhanced by the growth in GPU(CPU), IOT, Big Data, AI and other technologies have been promulgated to above field. This will supplement the security monitoring, process thus promoting the evolution of security monitoring system to the direction of intelligent AI as the core of a new generation of intelligent video monitoring system. This is indeed a sought-after solution needed in present time. Furthermore, video cloud— a new generation of intelligent video monitoring architecture focused using AI and big data are emerging to provide better feasible solutions. A snapshot of the hardware needed for such an IT infrastructure is shown in Figure (8). So, in nutshell, a cloud based architect is a '*must*' for today's requirements.

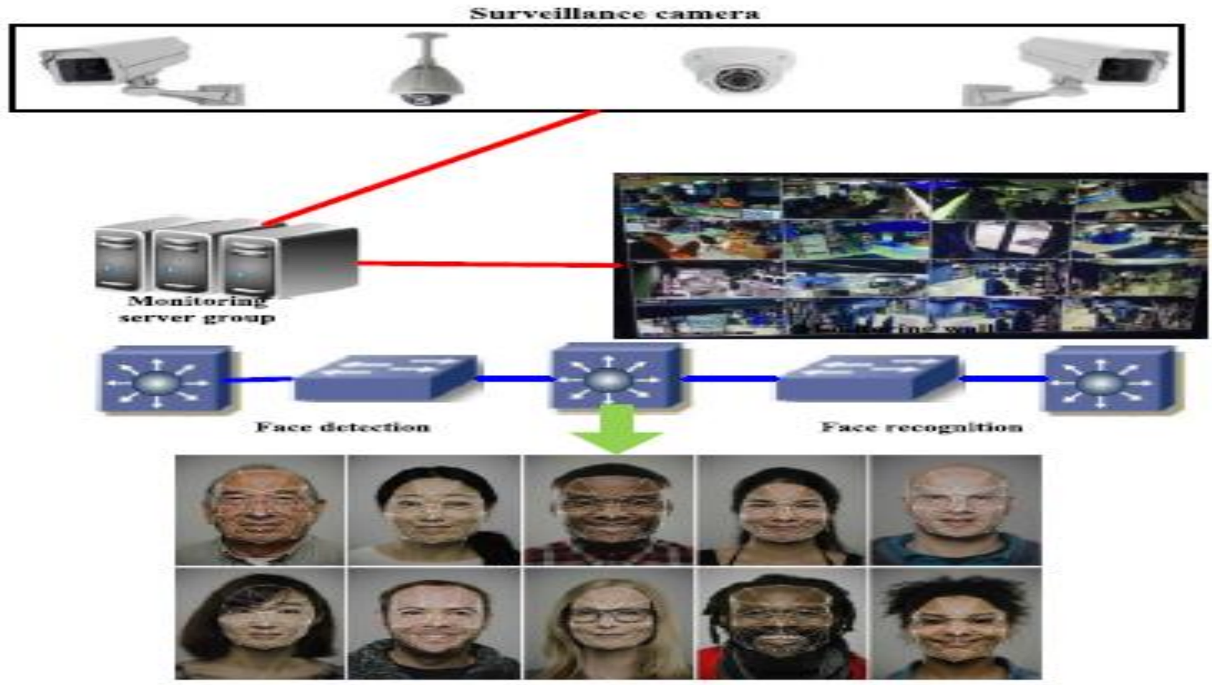


Figure 10 Images with differ rent facial expressions an pixel values

(Source: DOI: 10.1109/ACCESS.2020.2982779)

4.1 Video Surveillance

The main purpose of video surveillance retrieval is to locate the cause of an event and the development process of its association. The key information data of an event includes. The richer the information transferred by the search conditions, the more accurate the location and the simpler the search algorithm. On the contrary, if the information about the location if fuzzy, it would be difficult for search algorithms to pin-pointedly locate the case-in-hand. On the other side, a easy retrieved is expected to locate accurately in time. Further, the recent advancement is retrieval technology has offer greater opportunities retrieving monitoring and onerousness. It uses video segmentation, automatic digitization, voice recognition, shot detection, key frame extraction, automatic content association, video structure and other technologies, based on the knowledge of image processing, pattern recognition, computer vision, image understanding and other fields [37,38]. Through automatic intelligent analysis and preprocessing, the disordered and illogical monitoring video content (pedestrians, vehicles, and moving objects) is combined, and the key information of events and targets in the video is automatically obtained, and the video content and index are generated based on these information. In order to improve the computing speed, the cluster mode is adopted at present, which can provide real-time fast analysis ability, speedy and can be expanded according to the application needs, improve the computing ability, and save the case handling time. Fundamentally, the application of video retrieval technology in security monitoring is based on intelligent video analysis technology.

This refers to the use of computer image visual analysis technology, by separating the background and target in the scene and then analyzing and tracking the target in the camera scene. In current years, big data based predictive analytics are being used for video storage and retrieval. Formally, now video streaming is the technology for speedy retrieval and provides better efficiency on cloud based platforms. A detailed discussion on this can be found in [37, 38]

5. CONCLUSION AND FURTHER SCOPE

With the continuous development of social economy and the demand of market economy, more and more enterprises are gradually aiming at face recognition products of surveillance video, and the article has high-lighted the artifacts of face recognition system and shed the light on various perspectives. The mathematical derivation is kept minimum in order to have glance of the terms used in the concepts. Moreover, many textbooks and articles cover. Mathematical description in wider range. As people live in society with market- driven economy, the need of face recognition have acquired paramount place of importance, together with surveillance. face recognition products are everywhere like current traffic and housing construction and the video monitoring system in the field of public security focuses on the pre-warning analysis of the collected video images, however the post video analysis takes a lot of manpower and energy. Therefore, face recognition technology has become a subtle need in the field of public security video monitoring, which reduces the probability of illegal crime and maintain the stability in the society and across country.

The article covers some aspects of deep learning detection network for video single frame face detection. By tuning of hyper-parameter in hidden layers, layer by layer network performs better and improve the calculation efficiency, and using skin color detection to generate candidate areas in the preprocessing layer. Face recognition is the key of biometric recognition, and popular research front of AI is face recognition technology. With the explorative of technology, the application of face recognition technology in intelligent video surveillance system has broad application prospects and great practical significance.

In this article an attempt is being made towards face recognition and face detection with contemporary deep learning technology. The advancement has offered both merits and limitations however, with the ever increasing demands of face recognition, it is expected that in coming years, the growth will have experimental rising in recognizing facial systems and an increased securities surveillance with less vulnerabilities to threat. The use of CNN along with robust algorithms have been cited in the literature and these promising technologies would pave a subtle role in the coming years for better intelligent systems having secured surveillance and will thus provide a better quality life with minimum hassles. Furthermore, the promising GAN technology has already laid firm foundation towards facial systems and surly it would make the human-race less counterproductive instead!

KEY NOTES

(1) https://en.wikipedia.org/wiki/Gabor_filter

(2) https://en.wikipedia.org/wiki/Restricted_Boltzmann_machine

REFERENCES

[1] Hiremath, P.S., D. Ajit and C.J. Prabhakar, 2007. Modelling Uncertainty in Representation of Facial Features for Face Recognition I-Tech. Vienna, Austria, pp: 558. ISBN: 978-3-902613-03-5.

[2] Khanfir, S. and Y.B. Jemaa, 2006. Automatic facial features extraction for face recognition by neural networks, 3rd International Symposium on Image/Video Communications over fixed and mobile networks (ISIVC),Tunisia.

[3] Wiskott, L., J.M. Fellous, N. Kruger and C.V.D. Malsburg, 1999. Face Recognition by Elastic Bunch Graph Matching. Intelligent Biometric Techniques in Fingerprint and Face Recognition, Chapter 11, pp: 355- 396.

[4] Turk, M. and A. Pentland, 1991. Eigenfaces for recognition. J. Cogn. Neurosci., 3(1): 71-86.

[5] Hsu, R.L., M.A. Mottaleb and A.K. Jain, 2002. Face detection in colour images. IEEE T. Pattern Anal., 24(5):696-706.

[6] Basavaraj, A. and P. Nagraj, 2006. The Facial Features Extraction for face Recognition based On Geometrical approach. Canadian Conference on Electrical and Computer Engineering, CCECE' 06, P.D.A. College of Engineering, Gulbarga, pp: 1936-1939.

[7] Zhao, W., R. Chellappa, P.J. Phillips and A. Rosenfeld, 2003. Face recognition: A literature survey. ACM Comput.Surv., 35(4): 399-458.

[8] Zhang, H., B. Zhang, W. Huang and Q. Tian, 2005. Gabor wavelet associative memory for face recognition.IEEE T. Neural.Networ., 16(1).

[9] Eickler, S., S. Mu Ller and G. Rigoll, 2000. Recognition of JPEG compressed face images based on statistical methods. Image Vis. Comput., 18(4): 279-287.

[10] Hiremath, P.S. and D. Ajit, 2004. Optimized geometrical feature vector for face recognition. Proceedings of the International Conference on Human Machine Interface, Indian Institute of Science, Tata McGraw- Hill, Bangalore, pp: 309-320, (ISBN: 0 07- 059757- X).

[11] Belhumeur, V., J. Hespanha and D. Kriegman, 1997. Eigenfaces vs. Fisherfaces: Recognition using class specic linear projection. IEEE T. Pattern. Anal., 19(7): 711-720.

[12] Yousra, B.J. and K Sana, 2008. Automatic gabor features extraction for face recognition using neural networks, IEEE 3rd International Symposium on Image/Video Communications over fixed and Mobile Networks (ISIVC).

[13] Ranu Gorai (2016), A Survey of Digital Image Processing,International Journal of Research in Engineering, Technology and Science, VI(Special Issue)

[14] JangalaSasiKiran, N Vijaya Kumar, N SashiPrabha and M Kavya (2015), A Literature Survey son Digital Image Processing Techniques in Character Recognition of Indian Languages, International Journal of Computer Science and Information Technologies, 6(3), 2065-2069 .

[15] B.Thamotharan, M.Menaka, SandhyaVaidyanathan,SowmyaRavikumar (2012), Survey on image processing in the field of de-noising techniques and edge detection techniques on radiographic images,Journal of Theoretical and Applied Information Technology, 41(1), 26-34.

[16] Shafiqua T. Pathan, Sharyu S. Waghmare, Pranali P. Khadse, Pratik P. Shukla and Shrikant A. Wawre (2014), A Survey Paper On A Novel Approach For Image Classification Based On SUSAN Low Level Image Processing Algorithm From Real Time Video, International journal of scientific & technology research,3(2)

[17] DipenSaini (2016), Assembling of Human Beings in an Image to Detect a Group on the basis of Distance using Digital Image Processing, International Journal of Current Engineering and Technology, 6(2), 462-466

[18] MostafaBayati and Reza Fotouhi (2018), A Mobile Robotic Platform for Crop Monitoring, Advances in Robotics, 7(1), 186-191

[19] BahriHaythem, SayadiFatma, ChoucheneMarwa, Hallek Mohamed and Atri Mohamed (2016), Accelerating Fourier Descriptor for Image Recognition Using GPU, Applied Mathematics & Information Sciences An International Journal, 10(1), 297-306

[20] Zijiang Zhu, Hang Chen, Yi Hu and Junshan Li (2018), Age estimation algorithm of facial images based on multi-label sorting, EURASIP Journal on Image and Video Processing,114

[21] L. Sirovitch and M. Kirby, “Low-Dimensional Procedure for the Characterization of Human Faces,” J. Optical Soc. of Am.A, vol. 2, pp. 519-524, 1987.

[22] M. Turk and A. Pentland, “Eigenfaces for Recognition,” J. Cognitive Neuroscience, vol. 3, no. 1, 1991.

[23] M. Turk and A. Pentland, “Face Recognition Using Eigenfaces,” Proc. IEEE Conf. on Computer Vision and Pattern Recognition, 1991,pp. 586-591.

[24] P. Hallinan, “A Low-Dimensional Representation of Human Faces for Arbitrary Lighting Conditions,” Proc. IEEE Conf. Computer Vision and Pattern Recognition, 1994, pp. 995-999.

[25] H. Murase and S. Nayar, “Visual Learning and Recognition of 3-D Objects from Appearance,” Int’l J. Computer Vision, vol. 14, pp. 5-24, 1995.

[26] C. Lui and H. Wechsler, Independent component Analysis of Gabor features for face recognition, IEEE Trans on Neural Networks, Vol.14, No.4, July 2003.

[27] T.S.Lee, Image representation using 2D Gabor wavelets, IEEE trans. On Pattern Analysis and Machine Intelligence, Vol.18, No10, pp 859-970, 1996

[28] H. Zhang, B. Zhang, W. Huang and Q. tian, Gabor wavelet associative memory for face recognition, IEEE Trans on Neural Networks, Vol.16, No.1, January 2005.

[29] B. L. Zhang, H. Zhang and S. S. Ge, Face recognition by applying Wavelet sub band Representation and kernel associative memory, IEEE Trans on Neural Networks, Vol.15, No.1, January 2004.

[30] L. Wiskott, J. M. Fellous, N. Kruger and C. V. D. Malsburg, Face Recognition by Elastic Bunch Graph Matching, Intelligent Biometric Techniques in Fingerprint and Face Recognition, Chapter 11, pp. 355- 396, 1999.

[31] P. SmoIensky. Information processing in dynamical systems: Foundations of harmony theory.In D. E. Rumelhart and J. L. McClelland, editors, Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Volume 1: Foundations. MIT Press, 1986.

[32] J. Pearl. Probabilistic reasoning in intelligent Systems: networks of plausible inference. Morgan Kaufmann Publishers, San Mateo CA, 1988.

[33] S. Scardapane, M. Scarpiniti, and M. Bucciarelli, ``Microphone array based classication for security monitoring in unstructured environments," AEU- Int. J. Electron. Commun., vol. 69, no. 11, pp. 17151723, 2015.

[34] W. Glauser, ``Articial intelligence, automation and the future of nursing," Can. Nurse, vol. 113, no. 3, pp. 2426, 2017.

[35] B. Fan and R. Zhang, `Unmanned aircraft system and artificial intelligence," Wuhan DaxueXuebao (XinxiKexue Ban)/Geomatics Inf. Sci. Wuhan Univ., vol. 42, no. 11, pp. 15231529, 2017.

[36] D. Luo, G. Wen, D. Li, Y. Hu, and E. Huan, ``Deep-learning-based face detection using iterative bounding-box regression," Multimedia Tools Appl., vol. 77, no. 19, pp. 2466324680, Oct. 2018.

[37] Y. Zhao, F. Tang,W. Dong, F. Huang, and X. Zhang, ``Joint face alignment and segmentation via deep multi-task learning," Multimedia Tools Appl., vol. 78, no. 10, pp. 1313113148, May 2019.

[38] Y. Sun, X. Wang, and X. Tang, ``Hybrid deep learning for face verification," IEEE Trans. Pattern Anal. Mach. Intell., vol. 38, no. 10, pp. 19972009, Oct. 2016.