© 2019 IJSRSET | Volume 6 | Issue 2 | Print ISSN: 2395-1990 | Online ISSN : 2394-4099
Themed Section : Engineering and Technology
DOI : https://doi.org/10.32628/IJSRSET1962140

Video Face Recognition System for Large Scale Person Re-Identification Using Grassman Algorithm

T. Pandeeshvari[1], Aanjan Kumar[2]

[1]PG Student, Department of Computer Science and Engineering, Sri Raaja Raajan College of Engineering and Technology, Amaravathipudhur, Karaikudi, Tamil Nadu, India

[2]Assistant Professor, Department of Computer Science and Engineering, Sri Raaja Raajan College of Engineering and Technology, Amaravathipudhur, Karaikudi, Tamil Nadu, India

ABSTRACT

The identity or verification of humans primarily based on their thermal information isn't always an easy mission to perform, but thermal face biometrics can make contributions to that undertaking. Face reputation is an interesting and a successful application of Image analysis and Pattern recognition. Facial pictures are important for intelligent vision based human machine interaction. Face processing is based at the fact that the records approximately a consumer's identity may be extracted from the image and the computers can act as a consequence. A thermal face image should be represented with biometrics features that highlight thermal face characteristic and are compact and easy to use for classification. Second, image resolution is basically lower for video sequences. If the subject is present in very far from the camera, the actual face image resolution can be as low as 64 by 64 pixels. Finally, face image variations, such as illumination, expression, pose, occlusion, and motion, are more important in video sequences. The approach can address the unbalanced distributions between still images and videos in a robust way by generating multiple "bridges" to connect the still images and video frames. So in this project, implement still to video matching approach to match the images with videos using Grassmann manifold learning approach to know unknown matches. Finally provide voice alert at the time unknown matching in real time environments. And implement neural network classification algorithms to classify the face images in real time captured videos.

Keywords : Video Recognition, Thermal Face Detection, Facial Feature Extraction, Grassmann algorithm, Alert System

I. INTRODUCTION

In imaging technology, image processing is processing of pixel using mathematical operations with the aid of using any shape of signal processing for which the input is an image, a sequence of images, or a video, along with a image and audio; the output of image processing may be either an image or a hard and fast of characteristics or parameters related to the image. Most image-processing techniques involve treating the image as a two-dimensional sign and making use of popular sign-processing techniques to it. Images are also processed as three-dimensional alerts with the third-size being time or the z-axis. Image processing usually refers to virtual image processing, however optical and analog image graph processing is also possible. The acquisition of images (generating the enter image within the first region) is known as imaging.

Closely related to image processing are laptop images and computer vision. In computer pictures, images are manually crafted from bodily models of objects, environments, and lights, as opposed to being acquired (via imaging gadgets which includes cameras) from natural scenes, as in maximum animated films. In contemporary sciences and technologies, images also advantage much broader scopes because of the ever developing significance of medical visualization (of regularly massive-scale complex clinical/experimental records). Examples consist of microarray data in genetic research, or actual-time multi-asset portfolio trading in finance. Image evaluation responsibilities can be as simple as studying bar coded tags or as sophisticated as figuring out a person from their face.

Computers are indispensable for the evaluation of huge amounts of statistics, for duties that require complex computation, or for the extraction of quantitative records. On the other hand, the human visible cortex is superbimage graph analysis equipment, in particular for extracting higher-stage information and for lots packages which includes medicinal drug, safety, and faraway sensing human analysts still cannot get replaced with the aid of computers. For this cause, many critical image graph analysis gear such as area detectors and neural networks are stimulated with the aid of human visual perception fashions.

Many graphics based applications are capable of merging one or more images into a single file. The orientation and placement of each image can be controlled to improve efficiency. When selecting a raster image that is not rectangular, it needs identification and separation of the edges from the background, this is also known as silhouetting. This is the virtual analog of cutting out the photograph from a physical photograph. Clipping paths can be used to feature silhouetted images to vector portraits or web page format files that retain vector information. Alpha compositing lets in for gentle translucent edges when deciding on snap shots. There are some of methods to silhouette an picture with tender edges, which includes deciding on the picture or its heritage by way of sampling comparable colors, choosing the edges by raster tracing, or converting a clipping route to a raster choice. Once the photograph is chosen, it may be copied and pasted into some other phase of the identical record, or into a separate report. The choice may also be saved in what's referred to as an alpha channel. A popular manner to create a composite photo is to use obvious layers. The history image is used as the bottom layer, and the image with parts to be delivered are placed in a layer above that. Using an picture layer mask, all but the elements to be merged are hidden from the layer, giving the affect that those components have been added to the background layer. Performing a merge in this way preserves all of the pixel facts on each layers to extra easily enable destiny modifications in the new merged image.

Video Processing:

Video signal is essentially any collection of time various images. A nonetheless image is a spatial distribution of intensities that remain consistent with time, while a time various image has a spatial intensity distribution that varies with time. Video signal is dealt with as a series of images referred to as frames. An illusion of non-stop video is acquired by means of converting the frames in a quicker manner that is usually termed as frame rate. The demand for digital video is growing in regions including video teleconferencing, multimedia authoring systems, training, and video-on-demand systems.

Spatial Sampling:

The sensitivity of Human Visual System (HVS) varies according to the spatial frequency of an image. In the digital illustration of the picture, the value of every pixel desires to be quantized using some finite precision. In exercise, 8 bits are used in keeping with luminance pattern.

Temporal sampling:

A video consists of a sequence of images, displayed in rapid succession, to provide an illusion of non-stop movement. If the time gap between successive frames is too large, the viewer will take a look at jerky movement. The sensitivity of HVS drops off extensively at excessive frame rate. In exercise, maximum video codes use temporal sampling costs of 24 frames consistent with second and above. Video codes Digital video consists of video frames which are displayed at a prescribed frame rate. The basic frame rate is 30 frames per second. The frame layout specifies the size of character frames in phrases of pixels. The Common Intermediate Format (CIF) has 352 x 288 pixels, and the Quarter CIF (QCIF) layout has 176 x 144 pixels.

Frame types:

Three styles of video frames are I-frame, P-frame and B-frame. 'I' referred for Intra coded frame, 'P' stands for Predictive frame and 'B' stands for Bidirectional predictive frame. 'I' frames are encoded without any motion compensation and are used as a reference for future anticipated 'P' and 'B' kind frames. 'I' frames but require a enormously large number of bits for encoding. 'P' frames are encoded the usage of movement compensated prediction from a reference frame which can be both 'I' or 'P' frame. 'P' frames are extra efficient in terms of wide variety of bits required in comparison to 'I' frames, however still require more bits than 'B' frames. 'B' frames require the lowest variety of bits compared to both 'I' and 'P' frames however incur computational complexity. Frames between two successive 'I' frames, consisting of the leading 'I' frame, are collectively known as as Group of Pictures (GOP). It has one 'I' frame, two 'P' frames and six 'B' frames. Typically, more than one 'B' frames are inserted between two consecutive 'P' or between 'I' and 'P' frames. The lifestyles of GOPs helps the implementation of functions such as random get admission to, rapid forward or fast and everyday opposite playback. Video processing era has revolutionized the world of multimedia with products including Digital Versatile Disk (DVD), the Digital Satellite System (DSS), High Definition TV (HDTV), virtual nevertheless and video cameras. The exclusive regions of video processing include (i) Video Compression (ii) Video Indexing (iii) Video Segmentation (iv) Video monitoring and many others.

Video Indexing:

Video indexing is necessary to facilitate efficient content-based retrieval and browsing of visual information stored in large multimedia databases. To create an efficient index, a set of representative key frames are selected which capture and encapsulate the entire video content.

Subsampling:

The simple idea of subsampling is to reduce the size of the input video (horizontal measurement and / or vertical size) and for that reason the range of pixels to be coded previous to encoding system. At the receiver side decoded images are interpolated for display. This method may be considered as one among most primary compression techniques which also uses particular physiological traits of the human eye and for that reason gets rid of subjective redundancy contained inside the video information. This concept is also used to discover subjective redundancies contained in chrominance statistics, i.e., human eye is more touchy to modifications in brightness than to chromaticity changes. RGB format is not favored due to the fact R, G, B components are correlated and transmitting R,G,B components one after the other is redundant. To triumph over this, the input photo is split into YUV components (one luminance and chrominance components). Next, the chrominance components are subsampled relative to luminance element with a Y: U:V ratio particular to unique programs. Subsampling is denoted inside the layout X:X:X, where the primary digits represent the quantity of luminance samples, used as a reference and usually "4". The second and third digits are the wide variety of chrominance samples, with recognize to the quantity of Y samples.

Video compression performs an critical function in many virtual video applications consisting of virtual libraries, video on call for, and excessive definition television. A video sequence with body length of 176 X a hundred and forty four pixels at 30 frames in keeping with 2d and 24 bits in keeping with pixel would require 18.25 Mbps, making it impractical to transmit the video collection to transmit over standard phone traces in which statistics costs are commonly confined to 56,000 bits consistent with 2nd. This example illustrates the need for video compression. Effective video compression can be carried out by minimizing each spatial and temporal redundancy. A video consists of a series of frames in order to compress the video for efficientstorage and transmission, the temporal redundancy amongst adjacent frames should be exploited. Temporal redundancy means that adjoining frames are similar whereas spatial redundancy means that neighboring pixels are comparable. Video coding interprets video sequences into an efficient bitstream. This translation includes the removal of redundant statistics from video sequence. Video sequence contains styles of redundancies spatial and temporal. Removal of spatial redundancy is typically termed as interframe coding and removal of temporal redundancy is called as interframe coding. Video compression algorithms can be extensively labeled into kinds (i) Lossless video compression and (ii) Lossy video compression. Due to its significance in multimedia programs, most of the algorithms in video compression have focused on lossy video compression. Lossless video compression is vital to applications in Luminance sample (Y) Chrominance pattern (U, V) 7 which the video nice cannot tolerate any degradation including archiving of a video, compression of scientific and satellite television for pc motion pictures and so forth. Intra frame coding removing the spatial redundancy with a frame is normally termed as intraframe coding. The spatial redundancy inside a body is minimized with the aid of the use of remodel. The commonly used remodel is Discrete Cosine Transform. Inter frame coding the temporal redundancy among successive frames is removed by using interframe coding. Interframe coding exploits the interdependencies of video frames. Interframe coding relies at the fact that adjacent photographs in a video collection have excessive temporal correlation. To decrease the temporal correlation, a frame is chosen as a reference, and subsequent frames are anticipated from the reference.

II. RELATED WORK

Y. Yan, et al., [1] proposed a novel active sample selection approach (a.k.a. active learning) for image classification by using web images. Previous research has shown that cross-media modeling of various media types is beneficial for multimedia content analysis. The web images are often associated with rich textual descriptions (e.g., surrounding texts, captions, etc). While such text information is not available in testing images, we show that text features are useful for learning robust classifiers, enabling better active learning performance of image classification. Typical active sampling methods only deal with one media type which cannot simultaneously utilize different media types. The new supervised learning paradigm, namely learning using privileged information (LUPI), can be used to solve this problem. In a LUPI scenario, in addition to main features, there is also privileged information available in the training procedure. Privileged information can only be used in training, and is not available in testing. Uncertainty sampling is the most frequently used strategy in the active learning. In this work, propose to exploit both visual and text features for active sample selection by taking text as privileged information. By LUPI, train SVMs on visual features and slacking function on text features.

Y. Yang, et al., [2] proposed a new feature selection algorithm, which leverages the knowledge from related multiple tasks to improve the performance of feature selection. In this study, the following lessons have been learned: Sharing information among related tasks is beneficial for supervised learning. However, if the multiple tasks are not correlated, the performance is not necessarily improved. Compared to single task learning, the advantages of multitasks learning are usually more visible when only have few training examples per task. As increase the number of positive training data, the intra-task knowledge is sufficient for training, and thus adapting inter-task knowledge does not necessarily help. It is not always the case that feature selection improves the performance. However it is still beneficial because it improves the efficiency. Also, feature selection would provide us with better interpretability of the features. The improvement of feature selection varies when different classifiers are used. For example, since linear SVM actually has the ability to assign different weights to different features, the performance improvement of SVM is less than KNN, after feature selection. The optimization approach is proposed, followed by the proof of its convergence.

X. Chang, et al., [3]this proposed work aims to solve the limitations of the existing discriminant analysis algorithms for high-order data and propose a compound rank-k projection algorithm for discriminant bilinear analysis. Different from, the convergence of optimization approach is explicitly guaranteed. This adopts multiple orthogonal projection models to obtain more discriminant projection directions. In particular, here use h sets of projection matrices to find a low dimensional representation of the original data. The h projection matrices are orthogonal to each other. In this way, a larger search space is provided to find the optimal solution, which will yield better classification performance. Name the proposed algorithm as Compound Rank-k Projection for Bilinear Analysis (CRP). It is worthwhile noting that the algorithm can be readily extended to high-order tensor discriminant analysis. The main contributions of this work can be summarized as CRP can deal with matrix representations directly without converting them into vectors. Hence, spatial correlations within the original data can be preserved. The rest of this work is organized as summarizes an overview of the classical LDA as well as 2DLDA. A novel compound rank-k projection for bilinear analysis is proposed.

Y. Yang, et al., [4]proposed a framework consisting of two algorithms for multimedia content analysis and retrieval. First, a new transductive ranking algorithm, namely, ranking with Local Regression and Global Alignment (LRGA), is proposed. Differently from distance-based ranking methods, the distribution of the samples in the whole data set is exploited in LRGA. Compared with the inductive methods, only the query example is required. In contrast to the MR algorithm that directly adopts the Gaussian kernel to compute the Laplacian matrix , LRGA learns a Laplacian matrix for data ranking. For each data point, employ a local linear regression model to predict the ranking scores of its neighboring points. In order to assign an optimal ranking score to each data point, propose a unified objective function to globally align local linear regression models from all the data points. In retrieval applications, there is no ground truth to tune the parameters of ranking algorithms like MR. Therefore, it is meaningful to develop a new method that learns an optimal Laplacian matrix for data ranking. Second, propose a semi-supervised learning algorithm for long-term RF. A system log is constructed to record the history RF information marked by all of the users. Refine the vector representation of multimedia data according to the log information via a statistical approach. To that end, we convert the RF information into pairwise constraints, which are classified into two groups. The data pairs in the first group are semantically similar to each other, while the data pairs in the second group are dissimilar to each other. While LDA can be used to exploit these two types of information, the valuable information in the unlabeled data is not utilized.

Olutola Fagbolu, et al., [5]proposed a 3-fold method to secure re-identification. Firstly, endorse a filter out pairing neural network (FPNN) for character re-identity. This deep studying method has numerous vital strengths and novelties as compared with existing works. It collectively handles misalignment, imagemetric and geometric transforms, occlusions and background muddle beneath a unified deep neural community. During education, all of the key components are at the same time optimized. Each component maximizes its strength whilst cooperating with others. Instead of the usage of handcrafted capabilities, it routinely learns most beneficial capabilities for the challenge of man or woman re-identification from records, collectively with the gaining knowledge of imagemetric and geometric transforms. Two paired filters are implemented to different camera perspectives for feature extraction. The filter pairs encode imagemetric transforms. While present works anticipate move-view transforms to be uni-modal, the deep structure and its maxout grouping layer allow modeling a aggregate of complex transforms. Secondly, train the proposed neural community with carefully designed education techniques along with dropout, records augmentation, records balancing, and bootstrapping. These strategies deal with the troubles of misdetection of patch correspondence, overfitting, and excessive unbalance of nice and negative education samples in this venture. Thirdly, re-observe the person re-identification trouble and construct a massive scale dataset that could compare the effect brought via automatic pedestrian detection.

Existing Methodologies

The term multi-view face reputation, in a strict sense, only refers to conditions wherein a couple of cameras gather the subject (or scene) concurrently and an algorithm collaboratively utilizes the received images/motion pictures. But the time period has frequently been used to recognize faces across pose versions. This ambiguity does now not purpose any hassle for reputation with (nonetheless) pixels; a set of images simultaneously considering more than one camera and people involved in a single camera but at one of a kind view angles are equivalent as some distance as pose versions are worried. However, inside the case of video statistics, the two instances diverge. While a multi-digital camera gadget ensures the purchase of multi-view facts at any second, the risk of acquiring the equivalent records through the usage of a single digicam is unpredictable. Such differences turn out to be vital in non-cooperative recognition programs together with surveillance. For clarity, will name the more than one video sequences captured by synchronized cameras a multi-view video and the monocular video collection captured when the subject changes pose, a single-view video. With the superiority of digital camera networks, multi-view surveillance motion pictures have end up increasingly not unusual. Nonetheless, most current multi-view video face reputation algorithms take advantage of unmarried-view films. Given a pair of face photos to confirm, they look up inside the collection to "align" the face element's look in a single photograph to the same pose and illumination of the other photograph. This approach will also require the poses and illumination conditions to be predicted for both face snap shots. This "general reference set" idea has also been used to increase the holistic matching algorithm, in which the ranking of look-up consequences forms the premise of matching degree. There also are works which handles pose versions implicitly without estimating the pose explicitly.

III. FACE CLASSIFICATION USING GRASSMANN ALGORITHM

Face detection is the primary level of a face recognition system. A lot of research has been completed in this vicinity, maximum of that's efficient and powerful for still images simplest & could not be applied to video sequences without delay. Face recognition in videos is an present topic in the field of image processing, computer imaginative and prescient and biometrics over many years. Compared with still face popularity images comprise more abundant facts than a single image so video include spatio-temporal statistics. To enhance the accuracy of face popularity in movies to get more robust and strong recognition can be accomplished via fusing records of multi frames and temporal data and multi poses of faces in videos make it possible to explore form information of face and combined into the framework of face reputation. The video-based fully reputation has extra advantages over the photo-based fully reputation. First, the temporal statistics of faces can be applied to facilitate the recognition challenge. Secondly, greater effective representations, which include face version or superb-resolution images, can be acquired from the video series and used to improve reputation consequences. Finally, video- based fully recognition permits mastering or updating the situation version over the years to enhance popularity effects for future frames. So video based totally face recognition is also a totally hard hassle, which suffers from following nuisance elements inclusive of low satisfactory facial photos, scale variations, illumination adjustments, pose variations, Motion blur, and occlusions and so forth.

In the video scenes, human faces can have limitless orientations and positions, so its detection is of a spread of demanding situations to researchers. In current years, multi-camera networks have turn out to be more and more not unusual for biometric and surveillance structures. Multi view face reputation has come to be an energetic studies place in current years. In this work, a technique for video-based face reputation in digital camera networks is proposed. Traditional processes estimate the pose of the face explicitly. A strong function for multi-view reputation this is insensitive to pose versions is proposed in this task. The proposed function is advanced the use of the round harmonic illustration of the face, texture mapped onto a sphere. The texture map for the complete face is constructed by using again-projecting the photograph intensity values from every of the perspectives onto the surface of the round version. A particle filter is used to track the 3-D vicinity of the pinnacle the use of multi-view facts. Videos provide an automated and efficient way for characteristic extraction. In specific, self-occlusion of facial capabilities, because the pose varies, raises essential demanding situations to designing sturdy face recognition algorithms. A promising method to address pose variations and its inherent demanding situations is the use of multi-view statistics. In video based face reputation, awesome achievement has been made with the aid of representing movies as linear subspaces, which usually lie in a unique form of non-Euclidean space referred to as Grassmann manifold.

To leverage the kernel-primarily based strategies advanced for Euclidean area, several current strategies were proposed to embed the Grassmann manifold right into a high dimensional Hilbert area by using exploiting the well-set up Project Metric, which can approximate the Riemannian geometry of Grassmann manifold. Nevertheless, they unavoidably introduce the drawbacks from conventional kernel-based methods which include implicit map and excessive computational price to the Grassmann manifold. To triumph over such obstacles, advocate a singular approach to learn the Projection Metric at once on Grassmann manifold in place of in Hilbert space. From the angle of manifold gaining knowledge of, proposed technique can be appeared as performing a geometry-aware dimensionality discount from the unique Grassmann manifold to a lower-dimensional, greater discriminative Grassmann manifold in which extra favorable category may be done. And also provide neural community classification algorithm to categorise faces with progressed accuracy. Finally offer voice based totally alert machine with actual time implementation.

ALGORITHM

Grassmann algorithm:

Representing the facts on Grassmann manifolds is famous in some image and video recognition responsibilities. In unique, design complete rank mapping layers to convert input Grassmannian records into extra desirable ones, make the most orthogonal re-normalization layers to normalize the ensuing matrices, observe projection pooling layers to reduce the version complexity in the Grassmannian context, and devise projection mapping layers to show the ensuing Grassmannian information into Euclidean forms for ordinary output layers. To train the deep community, make the most a stochastic gradient descent placing on manifolds in which the connection weights are living on, and have a look at a matrix generalization of returned propagation to replace the established statistics. The famous packages of Grassmannian records inspire us to construct deep neural network architecture for Grassmannian representation studying. For this motive, the new community architecture is designed to take Grassmannian statistics at once as enter, and learns new favorable Grassmannian records which might be able to improve the final visual responsibilities. In other phrases, the new community pursuits to deeply examine Grassmannian facts on their underlying Riemannian manifolds in an stop-to-give up getting to know structure. To perform discriminant gaining knowledge of on Grassmann manifolds, many works embed the Grassmannian into a Euclidean space. This may be finished both by way of tangent space approximation of the underlying manifold, or with the aid of exploiting a high-quality particular kernel function to embed the manifold into a reproducing kernel Hilbert space. In each of such cases, any present Euclidean method can then be carried out to the embedded information, considering that Hilbert spaces respect Euclidean geometry. For example, first embeds the Grassmannian into a high dimensional Hilbert area, and then applies conventional Fisher analysis approach. Obviously, most of those techniques are restricted to the Mercer kernels and consequently constrained to apply best kernel primarily based classifiers. Moreover, their computational complexity increases steeply with the range of education samples.

The Grassmann manifold G(m, D) is the set of m-dimensional linear subspaces of the R and D. The G(m, D) is a m(D–m)-dimensional compact Riemannian manifold.

An element of G(m, D) can be represented by an orthonormal matrix Y of size D by m such that Y = Im, where Im is the m by m identity matrix. For example, Y can be the m basis vectors of a set of pictures in R D. However, the matrix representation of a point in G(m, D) is not unique: two matrices Y1 and Y2 are considered the same if and only if span(Y1) = span(Y2), where span(Y) represends the subspace spanned by the column vectors of Y . Equivalently, span(Y1) = span(Y2) if and only if Y1R1 = Y2R2 for some R1, R2 ∈ O(m). With this understanding, here will often use the notation Y when this actually mean its equivalence class span(Y), and use Y1 = Y2 when mean span(Y1) = span(Y2), for simplicity.

Formally, the Riemannian distance between two subspaces is the length of the shortest geodesic connecting the two points on the Grassmann manifold. However, there is a more intuitive and computationally efficient way of defining the distances using the principal angles.

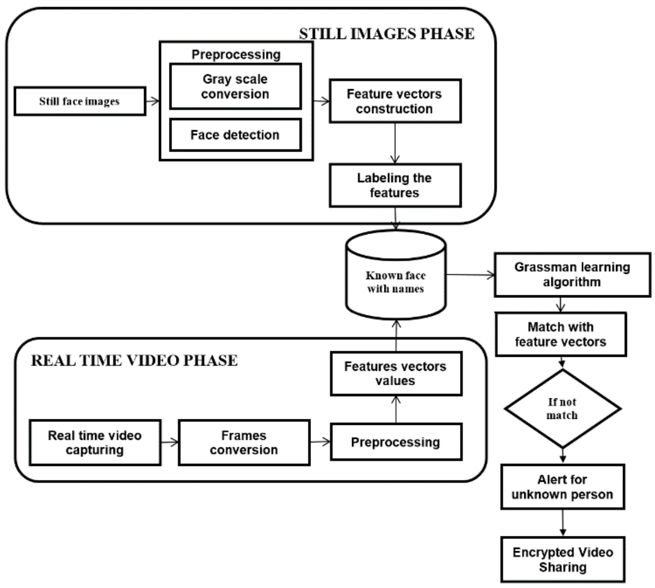

Fig 1. The proposed work

IV. EXPERIMENTAL RESULTS

Experimental results have the results of the proposed work. Proposed work was implemented using C#.NET is a front end and SQL is a back end process. This shows that the proposed work achieves higher security in home monitoring system with secure video sharing.

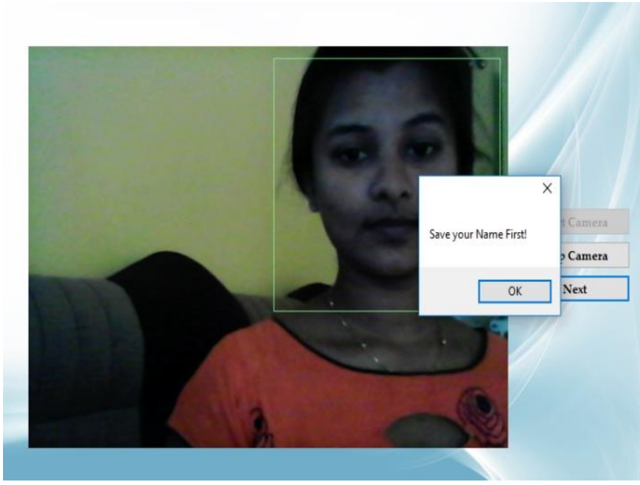

Fig 2 : Face Registration

Above figure shows the face registration process. Here user face image was captured using real time camera process. User should enter their name for registration. Facial features are extracted from image and stored on database with identification.

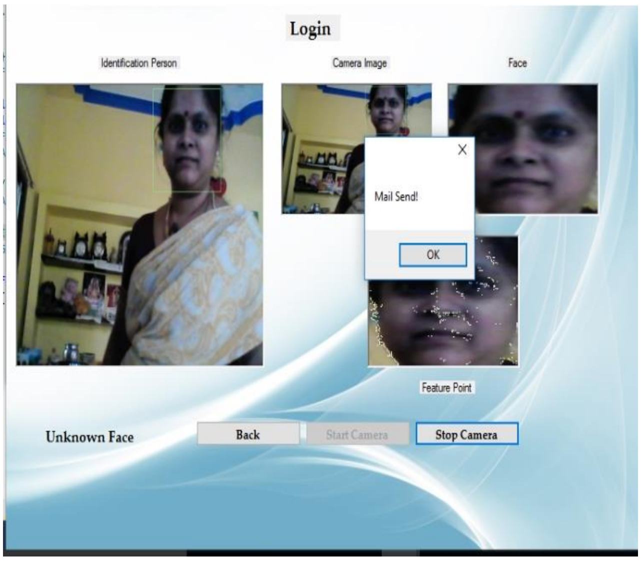

Fig 3 : Unknown Face Detection

Above figure shows the unknown face detection module. Features of the currently captured image are compared with database. When mismatch occurs on the features it will display the unknown face detection. Also send the alert message and unknown face image to the specified mobile number and email id.

V. CONCLUSION

In this project, reviewed face recognition technique for still images and video sequences. Most of these present approaches need nicely-aligned face images and handiest carry out either nevertheless picture face recognition or video-to video in shape. They are not suitable for face recognition under surveillance scenarios because of the following reasons: limitation in the number (around ten) of face images extracted from each video due to the large variation in pose and lighting change; no guarantee of the face image alignment resulted from the poor video quality, constraints in the resource for calculation influenced by the real time processing. Then proposed a local facial feature extraction based framework for still image and video-based face recognition under surveillance conditions. This framework is work on the basis of still-to-still, still-to-video and video-to video matching in real-time. While the training process uses static images, the recognition task is performed over video sequences. Proposed results show that higher recognition rates are obtained when use video sequences rather than statics – even when the algorithm using static images and that using video sequences address the same problems with exactly the same techniques. Evaluation of this approach is done for still image and video based face recognition on real time image datasets.

VI. REFERENCES

[1]. Y. Yan, F. Nie, W. Li, C. Gao, Y. Yang, and D. Xu, “Image classification by cross-media active learning with privileged information,” IEEE Transactions on Multimedia, vol. 18, no. 12, pp. 2494–2502, 2016

[2]. Y. Yang, Z. Ma, A. G. Hauptmann, and N. Sebe, “Feature selection for multimedia analysis by sharing information among multiple tasks,” IEEE Transactions on Multimedia, vol. 15, no. 3, pp. 661–669, 2013.

[3]. X. Chang, F. Nie, S.Wang, Y. Yang, X. Zhou, and C. Zhang, “Compound rank-k projections for bilinear analysis,” IEEE Transactions on Neural Networks and Learning Systems, vol. 27, no. 7, pp. 1502–1513, 2016.

[4]. Y. Yang, F. Nie, D. Xu, J. Luo, Y. Zhuang, and Y. Pan, “A multimedia retrieval framework based on semi-supervised ranking and relevance feedback,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 34, no. 4, pp. 723–742, 2012.

[5]. W. Li, R. Zhao, T. Xiao, and X. Wang, “Deepreid: Deep filter pairing neural network for person re-identification,” in Proc. CVPR, 2014, pp. 152–159.

[6]. S. Liao, Y. Hu, X. Zhu, and S. Z. Li, “Person re-identification by local maximal occurrence representation and metric learning,” in Proc. CVPR, 2015, pp. 2197–2206.

[7]. L. Zheng, Z. Bie, Y. Sun, J. Wang, S. Wang, C. Su, and Q. Tian, “Mars: A video benchmark for large-scale person re-identification,” in Proc. ECCV, 2016, pp. 868–884.

[8]. L. Zheng, L. Shen, L. Tian, S. Wang, J. Wang, and Q. Tian, “Scalable person re-identification: A benchmark,” in Proc. ICCV, 2015, pp. 1116–1124.

[9]. L. Zheng, S. Wang, L. Tian, F. He, Z. Liu, and Q. Tian, “Query-adaptive late fusion for image search and person re-identification,” in Proc. CVPR, 2015, pp. 1741–1750.

[10]. R. Xia, Y. Pan, H. Lai, C. Liu, and S. Yan, “Supervised hashing for image retrieval via image representation learning.” in Proc. AAAI, 2014, pp. 2156–2162.

Cite this article as :

Pandeeshvari.T, Aajan kumar, "Video Face Recognition System for Large Scale Person Re-Identification Using Grassman Algorithm", International Journal of Scientific Research in Science, Engineering and Technology (IJSRSET), Online ISSN : 2394-4099, Print ISSN : 2395-1990, Volume 6 Issue 2, pp. 550-559, March-April 2019. Available at doi : https://doi.org/10.32628/IJSRSET1962140

Journal URL : http://ijsrset.com/IJSRSET1962140