International Journal of Innovative Research in Technology & Science
ISSN: 2321-1156 Volume VIII, Issue IV, July 2020

Review of Face Recognition System

Md. Azheruddin[1], Sudhir Goswami[2], Shital Gupta[3]
CSE Department, SORT Peoples University
azhar.uddin100@gmail.com, sudhirpeople's01@gmail.com, email4sgupta@gmail.com

Abstract: *Face recognition is one of the most suitable applications of image analysis. It's a true challenge to build an automated system which equals human ability to recognize faces. While traditional face recognition is typically based on still images, face recognition from video sequences has become popular recently due to more abundant information than still images. Video-based face recognition has been one of the hot topics in the field of pattern recognition in the last few decades. This paper presents an overview of face recognition scenarios and video-based face recognition system architecture and various approaches are used in video-based face recognition system which can not only discover more space-time semantic information hidden in video face sequence, but also make full use of the high level semantic concepts and the intrinsic nonlinear structure information to extract discriminative manifold features. We also compare our algorithm with other algorithms on our own database.*

Keywords: Face recognition, image, video based face recognition

1. Introduction

Face recognition is a biometric approach that employs automated method to verify or recognize the identity of a living person based on his/her physiological characteristics. It also used in wide range of commercial and law enforcement and interesting area in real time applications. Face recognition has several advantages over other biometric technologies: It is natural, nonintrusive, and easy to use [1]. Face recognition system can help in many ways: for example some applications are Checking for criminal records and Detection of a criminal at public place, Finding lost children's by using the images received from the cameras fitted at some public places and detection of thief's at ATM machines, Knowing in advance if some unknown person is entering at the border checkpoints and so on. A face recognition system can operate in either or both of two modes: (1) face verification (or authentication), and (2) face identification (or recognition). Face verification involves a one to-one match that compares a query face image against a template face image. Face identification involves one-to-many matches that compare a query face image against all the template images in the database to determine the identity of the query face. The first automatic face recognition system was developed by Kanade[2], so the performance of face recognition systems has improved significantly.

Face recognition in videos is an active topic in the field of image processing, computer vision and biometrics over many years. Compared with still face recognition videos contain more abundant information than a single image so video contain spatio-temporal information. To improve the accuracy of face recognition in videos to get more robust and stable recognition can be achieved by fusing information of multi frames and temporal information and multi poses of faces in videos make it possible to explore shape information of face and combined into the framework of face recognition. The video-based recognition has more advantages over the image-based recognition. First, the temporal information of faces can be utilized to facilitate the recognition task. Secondly, more effective representations, such as a 3D face model or Super-resolution images, can be obtained from the video sequence and used to improve recognition results. Finally, video based recognition allows learning or updating the subject model over time to improve recognition results for future frames Face recognition can generally be categorized into one of the following three scenarios based on the characteristics of the Image to be matched. Such as Still-to-still recognition, Video-to-image face recognition, Video-to-video faces recognition [4].

i) Research on still image face recognition has been done for nearly half a century. Still-to-still image matching is the most common process and is used in both constrained and unconstrained applications. but it suffers from several factors those are the need to constrain the face recognition problem, computational constraints, and the large amount of legacy still face images (e.g. id cards, mug shots).

ii) Video-to-image face recognition can be seen as an extension of still image based face recognition. Video-to-still image matching occurs when a sequence of video frames is matched against a database of still images (e.g. mug shots or Identification photos). The input of the system is videos while the database is still face images. Compared to traditional still image based face recognition, how to explore the multi-frame information of the input video is the key to enhance the performance. In summary, image-video based methods make use of multi-frame information to improve the accuracy of face recognition, and improve the robustness to deal with pose variations, occlusions and illumination changes.

iii) Video-to-video matching, or re-identification, is performed to find all occurrences of a subject within a collection of video data. Re-identification is

generally a necessary pre-processing step before video-to-still image matching can be performed. Compared to video-image based methods, both the system input and the database this category are in the form of videos, which is a more difficult problem to solve. Based on the state of the arts, there are mainly three types of solutions this problem, those are Based on feature vector extracted from video input and Based on probability density function or manifold to depict the distribution of faces in videos and Based on generative models to describe dynamic variance of face in images.

2. RELATED WORK

By categorizing based on feature representation, recent methods in video-based face recognition (VFR) can be loosely organized into three categories: (1) direct modeling of temporal dynamics, (2) subspace-based representation, and (3) exemplar-based representation.
In video sequences, continuity is observed in both face movement and change in appearances.
Successful modeling of temporal continuity can provide an additional dimension into the representation of face appearances. As such, the smoothness of face movement can also be used for face tracking. Simultaneous tracking and recognition by Zhou and Chellappa is the first approach that systematically incorporates temporal dynamics in video-based face recognition (Zhou et al., 2003). A joint probability distribution of identity and head motion using sequential importance sampling (SIS) was modeled. In another tracking-and-recognition work (Lee et al., 2005), a nonlinear appearance manifold representing each training video was approximated as a set of linear sub-manifolds, and transition probabilities were learned to model the connectivity between sub-manifolds. Temporal dynamics within a video sequence can also be modeled over time using Hidden Markov Models (HMM) (Liu & Chen, 2003). Likelihood scores provided by the HMMs are then compared, and the identity of a test video is determined by its highest score. Due to the nature of these representations, many of these methods lack discriminating power due to disjointed person-specific learning. Moreover, the learning of temporal dynamics during both training and recognition tasks can be very time-consuming. Subspace-based methods represent entire sets of images as subspaces or manifolds, and are largely parametric in nature. Typically, these methods represent image sets using parametric distribution functions (PDF) followed by measuring the similarity between distributions. Both the Mutual Subspace Method (MSM) (Yamaguchi et al., 1998) and probabilistic modeling approaches (Shakhnarovich et al., 2002) utilize a single Gaussian distribution in face space while Arandjelovic et al. (Arandjelovic et al., 2005) extended this further using Gaussian mixture models. While it is known that these methods suffer from the difficulty of parameter estimation, their simplistic modeling of densities is also highly sensitive to conditions where training and test sets have weak statistical relationships. In a specific work on image sets subspaces using canonical correlations. Exemplar-based methods offer an alternative model-free method of representing image sets. This non-parametric approach has become increasingly popular in recent VFR literature. Krüeger and Zhou (Krüeger & Zhou, 2002) first proposed a method of selecting exemplars from face videos using radial basis function network. There are some comprehensive works (Fan & Yeung, 2006; Hadid & Peitikäinen, 2004) that proposed view-based schemes by applying clustering techniques to extract view-specific clusters in dimensionality-reduced space. Cluster centers are then selected as exemplars and a probabilistic voting strategy isused to classify new video sequences. Later exemplar-based works such as (Fan et al., 2005;Liu et al., 2006) performed classification using various Bayesian learning models to exploit the temporal continuity within video sequences. Liu et al. (Liu et al., 2006) also introduced a spatio-temporal embedding that learns temporally clustered key frames (or exemplars) which are then spatially embedded using nonparametric discriminate embedding. While all these methods have good strengths, none of these classification methods consider the varying influence of different exemplars with respect to their parent clusters.

3.VIDEO-BASED FACE RECOGNITION

Video based face recognition in image sequences has gained increased interest based primarily on the idea expressed by psycho physical studies that motion helps humans recognize faces, especially when spatial image quality is low. The traditional recognition algorithms are all based on static images but video-based face recognition has been an active research topic for decades. It is categorized into two approaches those are i) Set-based and ii) Sequential-based approaches[5]. Set-based approaches consider videos as unordered collections of images and take advantage of the multitude of observations where as sequence-based approaches explicitly use temporal information to increase efficiency or enable recognition in poor viewing conditions.

3.1 SYSTEM ARCHITECTURE

Video-based face recognition systems consist of three modules: i) Face detection module ii) Feature extraction module iii) Face recognition module.

3.1.1 Face detection

Face detection is the first stage of a face recognition system. This module system takes a frame of a video sequence and performs some image processing techniques on it in order to find locates candidate face region. System can operate on static images, where this procedure is called face localization and dealing with videos procedure is called face tracking. The purpose of face localizing and extracting the face region from the

background. Face detection can be performed based on several things those are skin texture, motion (for faces in videos), facial/head shape, facial appearance, or a combination of these parameters. An input image is scanned at all possible locations and scales by a sub window. Face detection is posed as classifying the pattern in the sub window as either face or non-face.

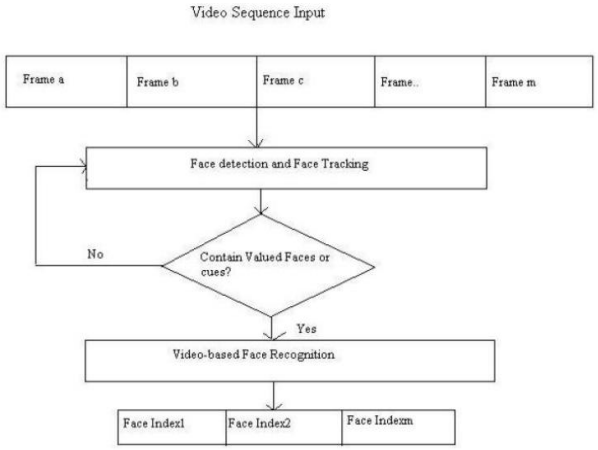

Fig.1 Process of face recognition in video

3.1.2 Feature extraction

The extraction of discriminant features is the most fundamental and important problem in face. After obtaining the image of a face, the next step is to extract facial features. There are two types of features can be extracted i) Geometric features ii) Appearance Features. Geometric features represent the shapes and location of facial components such as eyebrows, eyes, nose, mouth etc. Experimental results exhibited that the facial features cannot always be obtained reliably because the quality of images, illumination and other disturbing factors. The Appearance based features present the appearance (skin texture) changes of the face, such as Wrinkles and furrows.

3.1.3 Face recognition

Face recognition is the most significant stage in the entire system. Videos are capable of providing more abundant information than still image. The major advantages for using videos are Firstly the possibility of employing redundancy contained in the video sequence to improve still images recognition performance, second dynamic information is
available and thirdly to improve recognition effects from the video sequence using more effective representations such as a 3D face model or super-resolution images. Finally video-based recognition allows learning or updating the subject model over time .Though the advantages are obvious, there also exits some disadvantages. For example, poor video quality, low image resolution, and other influence factors (such as illumination, pose change, motion, occlusion, decoration, expression, large distance from camera, etc).The face recognition methods divided into two categories such as i) Frame-based recognition ii) Sequence-based recognition. The Frame-based recognition method is based on static images and sequence-based recognition method is based on dynamic video images. Sequence-based Expression recognition uses the temporal information of the sequence to recognize the expressions for one or more frames. Hidden Markov models (HMM), recurrent neural networks and rule based classifiers use sequence-based Expression Recognition. Sequence-based Expression Recognition classification schemes divided into two types such as dynamic and static classification. The static classifiers are classifiers that classify a frame in the video to one of the facial expression categories based on the tracking results of that frame. Mainly based on Bayesian network and Gaussian Tree-Augmented Naive (TAN) , Bayes classifiers. Dynamic classifiers are classifiers that take into account the temporal pattern in displaying facial expression. A multi-level HMM classifier is used for dynamic classification

4. HMM METHOD USED FOR VIDEO FACE RECOGNITION

Human face recognition is a subarea of object recognition which aims to identify a face given a scene or still images. It is very complex problem with high dimensionality due to the nature of digital images. Hidden Markov model to recognize human face from frames sequence. The proposed model trains HMM on the training data and then improves the recognition constantly using the test data. A sample figure is displayed in Figure 1 that captures the following:

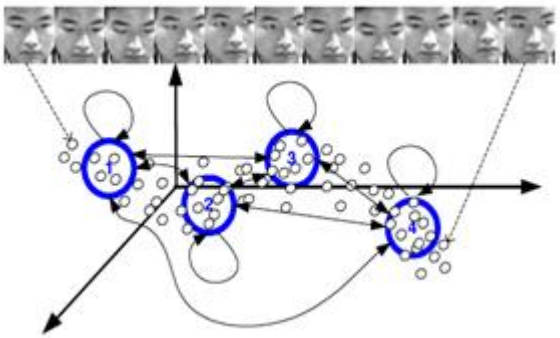

Fig.2: Temporal HMM graph

- HMM is used to study the temporal dynamics in the training process
- Then the temporal features of this test sequence is analyzed over time by the HMM of each subject
- The likelihood's are then compared to obtain the identity of the test video sequence

One advantage of this proposed idea is that the model can include dynamical characteristics.

Hidden Markov Model (HMM):

Hidden Markov Model is graphical model that suitable to represent sequential data. HMM consists of initial state π_i, unobserved states q_t, transition matrix A, and emission matrix B. HMM characterized by $\lambda = (A,B,\pi)$:

Given N of states $S = S_1,S_2,,S_N$ and q_t state of time T

A. A transition matrix where a_ij is the (i,j) entry in A:

$a_ij = P(q_t = S_j \mid q_{t-1} = S_i)$ where $1 \leq i, j \leq N$

B. the observation pdf $B = b_i(O)$

$$b_i(O) = \sum_{k=1}^{M} c_{ik} N(O, \mu_{ik}, U_{ik})$$

where $1 \leq i \leq N$

where c_{ik} is the mixture coefficient for k_th mixture component of S_i, M number of component in Gaussian mixture model .μ_{ik} is the mean vector and U_ik is the covariance matrix .the initial state $\pi_i = p(q_t = S_i)$ where $1 \leq i \leq N$

Conventional extensions to the basic Markov model are generally limited to increasing the memory of the system (durational modeling), which give the system compositional state in time. We are interested in systems that have compositional state in *space*, e.g., more than one simultaneous state variable. Recently, Jordan, Saul, and Ghahramani have developed a variety of multiple-HMM classifiers, including factorial HMMs [5] for independent processes; linked HMMs [8] that model non causal (contemporaneous) symmetrical influences; and hidden Markov decision trees [7] that feature a cascade of non causal influences from master to slave HMMs. The training algorithms are based on equivalence between HMMs and a class of Boltzmann machine architectures with tied weights [9, 10]. The linked HMM accepted, these algorithms use mean-field approximations from statistical mechanics. We present an algorithm for coupling HMMs with causal (temporal), possibly asymmetric influences. Theoretical and empirical arguments for this architecture's advantages can be found in [2]. To illustrate the difference between causal and non causal couplings, imagine modeling opponents in a tennis match: The non causal HMM couplings can represent the fact that it is unlikely to see both players playing net simultaneously; the causal HMM coupling can represent the fact that one player rushing to the net will drive the other back and restrict the kinds of returns he attempts. Here we introduce a coupling algorithm based on projections between component HMMs and a joint HMM; while performing the experiments described below we also perfected an algorithm with superior performance and lower complexity, based on an approximation to dynamic programming.

5. COUPLING AND FACTORING HMMS

Two HMMs are coupled by introducing table's conditional probabilities between their state variables. There is no simple decomposition of the prior probability that might lead to simple estimation procedures. The traditional workaround for modeling a system with two state variables forms a gross HMM from the Cartesian product of their states. This is unsatisfactory because the number of states is now squared and training data becomes very sparse on a per state basis. On the other hand, with a very large number of parameters it is very easy to raise the posterior probability of the model, but the result is gross over-fitting of the data and consequently poor generalization. Our algorithm takes this oversized parameter space and embeds within it a subspace manifold which represents all possible parameterizations of a much smaller system of coupled HMMs. Forward-backward analysis obtains posterior state probabilities in the larger space; we calculate the closest point on the manifold and re estimate so that the posterior probability of the model increases but the parameters stay on the manifold We obtain a joint HMM C from two component HMMs A,B by taking the Cartesian product of their states ai, bi and transition parameters $P_{a_i|a_j} P_{b_k|b_l}$. This results in a quadratic state table with joint states cij $c_{ij} = \{a_i, b_j\}$ We obtain transition and output probabilities as follows:

$$P_{c_{ik}|c_{jl}} = \Psi(P_{a_i|a_j}, P_{b_k|b_l}, P_{a_i|b_l}, P_{b_k|a_j}) \quad (1)$$

$$P_{c_{ik}}(o) = P_{a_i}(o) P_{b_k}(o) \quad (2)$$

Note that we have introduced coupling parameters P $P_{a_i|b_l} P_{b_k|a_j}$. If the composition functions Ψ is a Kronecker product, the following maximum-entropy factoring will factor project the joint HMM back into its components.

$$\hat{P}_{a_i|a_j} \propto \sqrt{\sum_l \sum_k P_{c_{ik}|c_{jl}}} \quad (3)$$

$$\hat{P}_{a_i} = \sum_k P_{c_{ik}} \quad (4)$$

These projections $(|\{A\}| \cdot |\{B\}|)^2$ factor the dimensional transition table of the joint HMM into $|\{A\}|^2 -$ and $|\{B\}|^2$ dimensional transition tables which parameterize two component HMMs. Note that we may just as easily define a projection which factors out the interaction between the component HMMs:

$$\hat{P}_{a_i|b_l} \propto \sqrt{\sum_j \sum_k P_{c_{ik}|c_{jl}}} \quad (5)$$

This is the basis of an algorithm in which a joint HMM is trained via standard HMM methods but constrained to factor consistently along both projections. As training increases its likelihood, we factor and reconstitute it, thus simultaneously training the component HMMs.

6. CONCLUSION

Hidden Markov models (HMMs) are used widely in perceptual computing as trainable, time-flexible classifiers of signals that originate from processes like speech and gesture. We believe that a conventional HMM is *not* a good model because most interesting signals fail to satisfy the restrictive Markov condition. Speech recognition researchers have grown increasingly frustrated with the performance of HMMs for this very reason, and vision researchers will run into it even faster. We have presented a mathematical framework for coupled hidden Markov models (CHMMs) which offers a way to model multiple interacting processes without running afoul of the Markov condition. CHMMs couple HMMs with temporal, asymmetric conditional probabilities.. In addition, CHMMs are far less sensitive to initial conditions than conventional HMMs, e.g., they are more reliable. We also compared CHMMs with linked HMMs (LHMMs), which have temporal, symmetric joint probabilities between chains. LHMM architecture shave been proposed as a desirable compositional HMM architecture.

References

[1]. R. Chellappa, C.L. Wilson, and S. Sirohey, "Human and machine recognition of faces: a survey", Proceedings of the IEEE, Vol.83, No.5, 1995, pp.705–741.

[2]. M. Turk and A. Pentland, "Eigen faces for Recognition", Journal of Cognitive Neuroscience, Vol.3, No.1, 1991, pp.71-86.

[3]. P.N. Belhumeur, J.P. Hespanha, and D.J. Kriegman, "Eigen faces vs. Fisher faces: Recognition Using Class Specific Linear Projection", IEEE Transaction on Pattern Analysis and Machine Intelligence, Vol.19, No.7, 1997, pp.711-720.

[4]. M. Lades, J.C. Vorbruggen, J. Buhmann, J. Lange, C. von der Malsburg, R.P. Wurtz, and W. Konen, "Distortion Invariant Object Recognition in the Dynamic Link Architecture", IEEE Transactions on Computers, Vol.42, No.3, 1992, pp.300-311.

[5]. Y. Li, Dynamic face models: construction and applications, PhD Thesis, Queen Mary, University of London, 2001.

[6]. G. J. Edwards, C.J. Taylor, T.F. Cootes, "Improving Identification Performance by Integrating Evidence from Sequences", In Proc. Of 1999 IEEE Conference on Computer Vision and Pattern Recognition, June 23-25, 1999 Fort Collins, Colorado, pp.486-491.

[7]. S. Zhou, V. Krueger, and R. Chellappa, "Face Recognition from Video: A CONDENSATION Approach", In Proc. of Fifth IEEE International Conference on Automatic Face and Gesture Recognition, Washington D.C., May 20-21, 2002, pp.221-228.

[8]. X. Liu, T. Chen and S. M. Thornton, "Eigen space Updating for Non-Stationary Process and Its Application to Face Recognition", To appear in Pattern Recognition, Special issue on Kernel and Subspace Methods for Computer Vision, September 2002.

[9]. A. Roy Chowdhury, R. Chellappa, R. Krishnamurthy and T. Vo, "3D Face Reconstruction from Video Using A Generic Model", In Proc. of Int. Conf. on Multimedia and Expo, Lausanne, Switzerland, August 26-29, 2002.

[10]. S. Baker and T. Kanade, "Limits on Super-Resolution and How to Break Them", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 24, No. 9, September 2002, pp.1167-1183.

[11]. L. Rabiner, "A tutorial on Hidden Markov Models and selected applications in speech recognition", Proceedings of the IEEE, Vol.77, No.2, 1989, pp.257-286.

[12]. A. Kale, A.N. Rajagopalan, N. Cuntoor and V. Krueger, "Gait-based Recognition of humans Using Continuous HMMs", In proceedings of the 5th IEEE International Conference on Automatic Face and Gesture Recognition, Washinton D.C. May 20-21, 2002, pp.336-341.

[13]. J.J. Lien, Automatic Recognition of Facial Expressions Using Hidden Markov Models and Estimation of Expression Intensity, doctoral dissertation, tech. report CMU-RI-TR-98-31, Robotics Institute, Carnegie Mellon University, and April 1998.

[14]. F. Samaria and S. Young, "HMM-based architecture for face identification", Image and vision computing, Vol.12, No.8, Oct 1994.

[15]. A. Nefian, A hidden Markov model-based approach for face detection and recognition, PhD thesis, Georgia Institute of Technology, Atlanta, GA. 1999.

[16]. J-L. Gauvain and C-H. Lee, "Maximum a Posteriori Estimation for Multivariate Gaussian Mixture Observations of Markov Chains", IEEE Transactions on Speech and Audio Processing, Vol.2, No.2, 1994, pp.291-298.

[17]. C. J. Leggetter and P. C. Woodland, "Maximum likelihood linear regression for speaker adaptation of the parameters of continuous density hidden markov models", Computer Speech and Language, Vol.9, 1995, pp. 171-185.

[18]. R. Gross and J. Shi, The CMU Motion of Body (MoBo) Database, tech. report CMU-RI-TR-01-18, Robotics Institute, Carnegie Mellon University, June, 2001.