A Review on Face Recognition and Video Databases

Sandeep Saxena¹, B Mohd Jabarullah², Dr. Mansaf Alam³

¹(Research Scholar, IFTM University, U.P, India)
²(Research Scholar, M. S. University, Tamil Nadu, India)
³(Jamia Millia Islamia, New Delhi)

Abstract: Each And Every Face Databases Have Its Own Limitations And Description To Test The Performance Of Face Recognition Technique. Even Though, Current Machine Recognition Systems Have Approached To Certain Level Of Recognition Rate. Their Success Is Limited By The Conditions Imposed By Many Real Applications And Constraint Imposed On Databases Such As Recognition Of Faces Acquired In An Outdoor Environment With Changes Of Illumination And Or Poses Remains A Largely Unsolved Problem. In View Of These This Paper Discusses Most Of The Video Databases And Face Recognition Methods. The Performance Of The Existing Methods On Video Datasets, NRC-IIT Video Data Set, The Honda/UCSD Video Data Set, Choke Point And MCGill Face Dataset Are Studied, Compare Their Performance In Terms Of Recognition Rate And Conclusion Reveals Future Development Of The Work.

Keywords: Video Database, Face Recognition

I. Introduction

Face Recognition Is An Important Research Area For Image Analysis And Understanding And Also In Addition To Numerous Applications Like Information Security(Application Security, Database Security, File Encryption, Internet Access, Medical Records), Law Enforcement And Surveillance(Video Surveillance, CCTV Control), Bioinformatics (Driving, License, Passports, Voter Registration, Aadhar Card Etc.), Smart Cards, Access Control Which Are Used For Essential Of Effective Communications And Interactions Among People. In Addition, The Problem Of Machine Recognition Of Human Faces Continues To Attract Researchers From Discipline Like Pattern Recognition, Neural Networks, Computer Vision, Computer Graphics, And Psychology. The Most Important Need For User Friendly Systems Which Is To Require That Can Secure Assets And Protect Privacy Without Losing Individual Identity In A Society. [1]

In General, The 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(T.Kanade,1973)[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. A General Statement Of The Problem Of Machine Recognition Of Faces Can Be Of Two Types One Is Of Still And Another One Is Video Image Identify Or Verify A Person From Stored Database Of Faces. In This Paper Provides Review Of Human Face Recognition Databases And Its Applications Till Date. This Is Also Need Of Wide Range Of Commercial And Law Enforcement Extraction From The Face Regions, Recognition, Or Verification.

In Order To Achieve Good Recognition Under Any Circumstances, The Images Captured By CCTV Are Pre Processed. One Of The Pre-Processing Process Is Noise Removal Because The Most Of The Images Are Affected By Noises While Capturing, Storing And Transmitting From One Resource To Another[3].

The Remaining Part Of The Paper Is Organized As Follows Section II Presents Analysis Of Face Video Image Databases Section III Describes Video Based Face Recognition Section IV Shows Experimental Studies, Section V Explains Conclusion And Future Work..

II. Analysis Of Face Video Image Databases

The Study Of Video Database Is Essential To Produce Efficient Mechanisms For Detection Of Faces Prior To Its Recognition And Identifying The Specifications, Limitations And Factors Which Impact On Recognition Rate Of Existing Databases. On Account Of This Study, Video Face Databases Such As INRC-IIT, Honda/UCSD, Choke Point, MCGill Are Considered.
2.1 NRC-IIT Video Database

To Recognize and Identify an object from video is very tedious because of video provides images of very low quality and also of low resolution. Further, faces in video are normally captured in unconstrained environment mostly under poor lighting, in motion and at a distance. By using neuro associative approach for recognition which can both learn and identify an object from low resolution low quality video sequences. This approach is derived from a mathematical model of biological visual memory, in which correlation-based projection learning is used to memorize. A face from video sequences and attractor-based association is performed to recognize a face over several video frames presented by Dmitry O. Gorodnichy et al. [4]. As per this study and observation, the complete specification and constraints of NRC-IIT Video Dataset is given as follows:

IIT-NRC Video-Based Facial Database was created with the aim to examine the computer’s ability to recognize faces in conditions known to be sufficient for humans, particularly the condition of low resolution close to the nominal face resolution of 12 pixels between the eyes. It contains pairs of 20-second 160*120 Mpeg-encoded video clips, each showing a face of computer user sitting in front of the monitor exhibiting a wide range of facial expressions and orientations as captured by a CMOS webcam mounted on the computer monitor. Video files of person faces in the database are very small due to small resolution and compression (less than 1 MB). This size is comparable to the size of ICAO-conformed high-resolution face images used to archive facial images for forensic purposes since video is often more informative than static. This database is most suited for testing the recognition performance with respect to such factors as (1) Low Resolution (2) Motion Blur (3) Out-Of Focus Factor (4) Facial Expression Variation (5) Facial Orientation Variation and (6) Occlusion without taking into account illumination changes.

![Fig. 1 Video Clip For 1 Sec.](image1)
![Fig. 2 Video Clip For 6 Sec](image2)
![Fig. 3 Video Clip For 9 Sec.](image3)
![Fig. 4 Video Clip For 13 Sec.](image4)

These video clips are taken from then IIT-NRC video dataset in which videoclips taken of low resolution within 1 sec in Fig. 1 and 6 sec in Fig. 2 with video blurring and 9 sec in the Fig. 3 videoclip is taken out of focus and at 13 sec in the Fig. 4 with changes in facial expression. It is observed, the following limitations are obtained from this study:
- Frontal Faces Of Different Subjects
- Background Is Static (Like Things Kept Are Stationary As Computer, Table, Almirah Etc.)
- Little Tilted Face With Time (Dynamic Position W.R.T Head Movement With Time)
- Images Are Taken Both Male And Female But In These >90% Images Are Taken From Gender Male And Rest Of From Female.
- View Of Subjects Are Taken Above The Shoulder.
- Subjects Are Taken Of The View (Neutral, Disgust, Laugh, Surprised, Sad Etc. Of Different Facial Expression)
- Age Of Subjects Are Around In Between 30 To 45.
2.2 The Honda/UCSD Video Database

This Video Database Consists Of Two Sets Of Data, The First Dataset Is Recorded By SONY EVI-D30 Camera At Honda Research Institute At 2002. It Includes Three Different Subsets, Training, Testing And Occlusion Testing. Each Subject Contains 20, 42, 13 Videos Respectively, Presented By Kaung-Chih Lee Et. Al. [5]. Second Dataset Is Recorded By SONY DFW-V500 Camera At Computer Vision Laboratory, University Of California, San Diego In 2004. It Includes Two Subsets Training And Testing Of 30 Videos From Another 15 Different Human Subjects, Presented By Kaung-Chih Lee Et. Al. [6].

Each Video Sequence Is 640*480 Resolution. All The Video Sequences Contain Significant 2-D (In-Plane) And 3-D (Out-Of-Plane) Head Rotations.

![Fig. 5 Subject Frontal Face](image1)
![Fig. 6 Subject Of Different Head Position](image2)

It Is Observed From This Database That Every Frame Of Video Sequence Is Recorded Both In An Indoor And Outdoor Environment With Static Background, At 15 Frames/Second.

2.3 Choke Point:

The Third Database, Choke Point Presented By Y. Wong Et. Al. [7], Which Is Specified For Person Identification/Verification Under Real-World Surveillance Conditions. An Array Of Three Camera Was Placed Above Several Portals To Capture Subjects Walking Through Each Portal In A Natural Way While A Person Is Walking Through A Portal, A Sequence Of Face Images Captured. This Video Set Is Useful For Person Re-Identification And Face Quality Measurement.

![Fig. 7 Camera 1](image3)
![Fig. 8 Camera2](image4)
![Fig. 9 Camera3](image5)
The Video Dataset According To This Study Reveals The Following Observations: Resolution Of Video Sequence Is 800*600 Pixels. An Array Of Three Camera Placed Above Several Portals At Equal Distance To Capture Subjects Walking Through Each Portal A Natural Way. One Of The Camera Is Likely To Capture A Face Set Where Where A Subset Of The Faces Is Near Frontal. The Data Set Consists Of 25 Subjects (>70% Of Male Subjects). The Dataset Has Frame Rate Of 30 Fps. In Total, The Toal Dataset Consists Of 48 Video Sequences And 64,204 Face Images. Only One Subject Is Presented In The Image At A Time In All The Sequences.

It Is Observed That In The Video Face Dataset, Only Three Cameras Has Been Used To Capture Subjects At Equal Distance, Frontal Views And Not Highlighting Object Of Interest.

### 2.4 McGill Faces Databases

The Study And Observation Of Fourth Dataset Reveals The Following: A Face Video Database Of Video Sequences Collected 18000 Video Frames Of 640*480 Resolution From 60 Video Sequences, Each Of Which Is Recorded From Equally Different Subject. Individual Subject Of Different Gender Is Taken Of The Frontal Face And The Head Position Is Rotated And Background Is Static For All. This Databases Real-World Face Video Database, Presented By M.Demirkus Et. Al. [8].

The Following Fig. 10 Illustrates The Subject Snapshots Of Video At 14 Sec. Appears Down While Fig. 11 Illustrates At 19 Sec With Rotated, In Fig. 12 View At 110 Sec And In Fig. 13 At 114 Sec At Tilted Faces.

It Is Observed That The Background Of The Particular Subjects Remain Same.

![Figure 10 Subject Of View 14 Sec](image1)

![Figure 11 Subject Of View 19 Sec](image2)

![Figure 12 Subject Of View 110 Sec](image3)

![Figure 13 Subject Of View 114 Sec](image4)

### III. Video Based Face Recognition

Video-Based Face Recognition Systems Consist Of Three Steps I) Face Detection Module Ii) Feature Extraction Module Iii) Face Recognition Module. In Identification Problems, The Input To The System Is An Unknown Face, And The System Reports Back The Determined Identity From A Database Of Known Individuals, Whereas In Verification Problems, Systems Needs To Confirm Or Reject The Claimed Identity Of The Input Face. The General Concept Of Video Based Face Recognition System As In J.Sunetha.Et.Al.[9], Frames Are Collected In Particular Interval Of Time And Face Detection And Tracking Is Applied To Locate Face In Given Video Images. After Locating The Face In Each Frame The Faces Are Indexed. Subsequently, These Indexed Faces Are Considered As Input For Further Stages Of Video Based Face Recognition System. The Following Section Describes Reviews Of Existing Method In Each Stage Given Above, As Depicted In The Following Fig. 14 Video Based Face Recognition System.
3.1 Face Detection

Face Detection Is The First Step 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 , 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. Face Detection Techniques Similar To Those Applied For Still Images Are Then Employed To Find The Exact Location Of Faces In The Current Frame, Thus Initiating Face And Facial Feature Tracking. Face Tracking Techniques Include Head Tracking, Where The Head Is Viewed As A Rigid Object Performing Translations And Rotations. While The Tracking Module Finds The Exact Position Of Facial Features In The Current Frame Based On An Estimate Of Face Or Feature Locations In The Previous Frame(S). Face Detection Based On Videos Consists Of Generally Three Main Processes. First One Is Frame Based Detection, In This Process, Lots Of Traditional Methods[For Still Images Can Be Introduced Such As Statistical Modeling Method, Neural Network Based Method, SVM-Based Method, HMM Method, BOOST Method And Color-Based Face Detection, And Many More . Second One Is Combination Of Detection And Tracking, This Says That Detecting Face In The First Frame And Then Tracking It Through The Whole Sequence. Since Detection And Tracking Are Independent. Third One Is, Instead Of Detecting Each Frame, Temporal Approach Exploits Temporal Relationships Between The Frames To Detect Multiple Human Faces In A Video Sequence.

Rowly Et. Al. [10] Presented A Paper To Detect A Face Using Neural Network. Input Of Images Upright, Frontal View In Grey Scale And Generate Output Ranging From 1 To -1 , Signifying The Presence Of A Face Or Absence Of Face. Author Does Not Included Dynamic Background And Different Lighting Condition. However This Method Outperforms With Acceptable Detection Rate On Database CMU.

Yasaman Heydarzadeh Et Al. [11] Proposed An Efficient Face Detection Method Using Adaboost And Facial Parts And Learning These By Adaboost Algorithm. Occluded Faces Has Been Considered For Experimental Studies. This Improved Method Of Viola Approach Occlusion 94.7% Detection Rate With False Positive Rate 2% AT&T Databases Consists Of Occluded Faces And Considered For Analysis.

L. Crena Et. Al. [12] Proposed In His Paper . Input Image Extract The Feature Vector Of Each Candidate Image Sub Window And Classify A Face Or Non Face By Trained Model, Author Consider Only Whole Face Image And Suggested To Increase The Detection Rate By Using Eyes,Nose Etc That Is Combination Of Both Global Face And Local Facial Features[13]. The Performance Of Proposed Method Produces Reasonable Outcome.

3.2 Face Recognition

David G. Lowve Et. Al. [15] Proposed A Method Distinctive Image Features From Scale-Invariant Keypoints. The Scale Of Invariant Feature Transform( SIFT) Feature Are Extracted From Images To Help In Reliable Matching Between Different Views Of Same Subject. The Extracted Features Are Maintained To Scale And Orientation And Distinctive Of The Image. The First Step Computes The Locations Of Potential Points In The Image By Detecting The Maxima And Minima Of A Set Of Difference Of Gaussian DOG Filters Different Scales All Over The Image. Then, These Locations Are Refined By Discarding Points Of Low Contrast. An Orientation Is Then Assigned To Each Key Point On Local Image Gradient, Transferred According To The Orientation Of The Key Point To Provide Orientation Invariance.


M.Saraswathi Et. Al. [17] Proposed A Method Evaluation Of PCA And LDA Technique For Face Recognition Using ORL Face Database, The Images Based Face Recognition Has Classified In To Appearance Based Face Recognition And Model Based Face Recognition. Model Based Approach Involve 2-D Or 3-D Models. Appearance Based Face Recognition Classified Into Linear Or Nonlinear Analysis. Linear Analysis Methods The Face Vectors Can Be Projected To The Basic Vectors. PCA Is A Linear Transformation Also Known As Karhunen-Lo-Eve Transformation Which Capture The Variance Of Input Data. PCA Is Used For Dimensionality Reduction It Also Remove Redundancy And Compress Data. Use Of LDA Is For Feature Extraction. LDA Is Also Linear Transformation Implicitly Finds Within And Outside Class Difference.

Kohir Et. Al. [18] Proposed A Method For Face Recognition Using DCT-HMM Approach, 2-D Coefficients Are Used Along With 1-D HMM For Feature Extraction. The Features Are Obtained By Using Sliding Square Window In A Raster Scan Over The Face Image From Left To Right And With A Predefined Overlap. 2-D DCT Coefficients Are Computed At Every Position Of The Window Over The Image. These Coefficient Are Used To Form An Observation Vector. In Order To Evaluate The Performance Of The System , The Proposed Method Is Used Using The ORL Database .According To The Result The Performance Of Recognition Rate Is Substantially High When The Percentage Of Overlap Increases.

IV. Experimental Studies

4.1 Face Detection

Face Detection Methods Such As Neural Network Applied On CMU Database Which Includes 1050 Faces With 507 Frontal Views, The Method Adaboost With Haar Like Features Applied On MIT+CMU Database Which Contains 130 Graylscale Image With 500 Frontal Faces, These Test Images On Variable In Scale, Illumination, Situation And Orientation But Nearly Vertical. The SVM With HOG Method Worked On AT&T Database Which Include 2385 Faces And 705 Non Faces Are Considered For Training And Testing . The Viola Jones Face Detection Method Performed On MIT+CMU Frontal Databases. This Database Consists Of 130 Images With 507 Labeled Frontal Faces.

The Comparative Study Of All These Methods As Per Their Results Of Face Detection Isillustrated In Following Table(1) And Corresponding Graph Representation Is Shown In Figure.

<table>
<thead>
<tr>
<th>S.NO.</th>
<th>Method</th>
<th>Database Used</th>
<th>Detection Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Neural Network</td>
<td>CMU</td>
<td>90.5%</td>
</tr>
<tr>
<td>2.</td>
<td>Adaboost With Haar Like Feature</td>
<td>MIT+CMU</td>
<td>94.7%</td>
</tr>
<tr>
<td>3.</td>
<td>SVM With H.O.G</td>
<td>AT&amp;T</td>
<td>85.71%</td>
</tr>
<tr>
<td>4.</td>
<td>Viola Jones</td>
<td>MIT+CMU</td>
<td>94.1%</td>
</tr>
</tbody>
</table>
As Per Above Table It Is Observed That The Maximum Detection Rate Is 94.7%. This Success Rate, Overall, Is Satisfying Certain Limitations Such As Illumination, Different Situation, Vertical Orientation Or Frontal Face Images. Hence, Each Of The Above Method Has Its Own Limitation And Advantages. Therefore With The Conclusion Of This Study Each Method Is Well Suited For Particular Circumstances And Database Used.

4.2 Face Recognition


The Comparative Study Of All These Methods As Per Their Observations Of Face Recognition Illustrated In Following Table(2) And Corresponding Graph Representation Shown In Figure(2).

<table>
<thead>
<tr>
<th>S.No.</th>
<th>Method</th>
<th>Database Used</th>
<th>Recognition Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>SIFT</td>
<td>Yale</td>
<td>90.1%</td>
</tr>
<tr>
<td>2</td>
<td>Eigenfaces</td>
<td>FRAV3D</td>
<td>96%</td>
</tr>
<tr>
<td>3</td>
<td>PCA+LDA</td>
<td>ORL</td>
<td>93.7%</td>
</tr>
<tr>
<td>4</td>
<td>HMM-DCT</td>
<td>ORL</td>
<td>99.5%</td>
</tr>
</tbody>
</table>

The Table Above Shows Analysis Result Of Four Different Methods. The HMM-DCT Method Gives Highest Rate Of Recognition In Comparisons Of Other Methods. However, The HMM-DCT Method Has Been Achieved Highest Recognition Rate Only In The Case Of 75% Of Overlapping Windows To Use To Extract DCT Feature. The In The Case Of 50% Overlapping Windows Used To Extract DCT Feature The Performance Is Drastically Decreased To 74%.
System With Fulfilling All Limitations. This Paper Highlights Some Methods And Available Methods Which Are Quite Generic And May Have Broader Applications. According To Reviewed Databases, The Most Of These Databases, In General, Considered For Evaluation Of Frontal Faces. Different Illumination Conditions, And Static Background And Indoor. Experiments On Complex Databases Are Difficult And Time Consuming. However, Improvement Of This Work Has To Include Images With Low Resolution And Images With Outdoor Dynamic Background. Similarly The Performance Evaluation Of Face Recognition Methods Revealed That HMM-DCT Methods Is Comparatively Much High Recognition Rate But This Method May Be Further Modified By Eliminating Limitations To Gain The Desired Output.

References


DOI: 10.9790/0661-1802031219 www.iosrjournals.org 19 | Page