International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

SMART ATTENDANCE MANAGEMENT SYSTEM

[1st] R.Ramani, [4th] K.Palpandi, [2nd] B.Manishankar , [3rd] J.Naga Chandhru

[1] Associate Professor, Dept of Computer science Engineering, P.S.R Engineering College, Sivakasi, Tamilnadu, rramani.ananth@gmail.com
[4] Dept of Computer science Engineering, P.S.R Engineering College, Sivakasi, Tamilnadu, palpandip049@gmail.com
[2] Dept of Computer science Engineering, P.S.R Engineering College, Sivakasi, Tamilnadu, mannishankarbpsrec@gmail.com
[3] Dept of Computer science Engineering, P.S.R Engineering College, Sivakasi, Tamilnadu, Nagachandru66@gmail.com

ABSTRACT-

Identifying a person captured on camera or in an image is one of the primary objectives of face recognition in surveillance scenarios. This means that faces in still photos and video clips must match. High-quality still picture automatic face identification can perform satisfactorily, while video-based face recognition is difficult to do at a comparable level. Video sequences have a number of drawbacks over still image face recognition. First off, the majority of the time, CCTV cameras provides low-quality photos. There is more background noise, and moving objects or out-of-focus subjects might cause photos to become blurry. Second, video sequences often have lesser picture resolution. In situations where the person is very far away from the camera, the resolution of the true facial image could be as low as 64 by 64 pixels. Lastly, in video sequences, differences in facial image, including lighting, emotion, position, occlusion, and motion, are more significant. By creating many "bridges" to link the still image and video frames, the method may effectively handle the uneven distributions between still photos and films. In order to match photos with videos and identify unknown matches, the Grassmann algorithm may be used in this research to develop a still-to-video matching strategy. Matching feature vectors based on deep learning techniques and reading the feature vectors using the Grassmann method. Additionally, when an unknown face is detected, send out an SMS and email notice. After that, provide reports for the attendance system.

KEYWORDS: Face recognition, Deep learning, Attendance system, Features extraction, Notification

1. INTRODUCTION

A method for identifying or verifying a person's identification based only on their face is called facial recognition. It can be used on images, videos, or situations that occur in real time. During police stops, law enforcement may employ facial recognition software on cell phones to identify people. Facial recognition software is not without flaws, though, particularly when it comes to accurately identifying young people, girls, and members of ethnic minorities like African Americans. Certain groups may be disproportionately impacted by this bias. When several cameras record pictures or videos, the term "multi-view face recognition" describes the scenario where an algorithm makes use of all the information gathered. Conversely, single-view face recognition concentrates on identifying faces in various orientations. It becomes important to distinguish between single-view and multi-view in non-cooperative surveillance systems. The prevalence of multi-view surveillance videos is rising along with the use of camera networks. Still, the majority of facial recognition algorithms are mostly concerned with single-view videos. These algorithms match up faces in many images with the same lighting and posture so that they can be compared. Matching algorithms have also employed the notion of a "ordinary reference set" to evaluate the degree of resemblance between faces.The standard configuration for face detection is seen in Figure 1.

2. RELATED WORK

M. Ayazoglu, B. Li, et.al,…[1] The suggested approach is predicated on the notion that a target's pathways ought to exhibit comparable patterns across various camera angles. This enables us to build a single model that, at any given time, explains all the data that is accessible from various cameras. The target's future location can then be predicted using this model. We can still estimate the target's location even if some cameras are unable to see it; we just need to make sure that the information from the other cameras matches our existing knowledge. This aids in our handling of scenarios in which the target is partially obscured from vision. The technique functions effectively even in the absence of camera calibration, 3D picture reconstruction, or sensor distance requirements. Experiments in the real world have demonstrated the effectiveness of this strategy even in the presence of moving or changing targets.

D. Baltieri, et.al,…[2] The study introduces articulated frame styles as a new 3D method for recognising the human body. This method maps appearance descriptors to skeletal bones using color, depth, and skeleton streams produced by the Microsoft Kinect sensor and OpenNi libraries. The generated signature is linked to the anatomical structure of the body, enabling characteristic-based descriptions and a computed metric that functions as characteristic selection and frame part weighting.

D. Baltieri,et.al,…[3] A tool for re-identifying humans based on appearance descriptors applied to three-dimensional frame styles is called the SARC3D system. In comparison to other computer vision disciplines, this is comparatively new. On the other hand, precise 3-D orientation estimation, segmentation, and identification are necessary. Partial issue mitigation with 2D models allows for part-based matching, however orientation-specific problems could occur.

I. B. Barbosa, et.al,…[4] The study addresses reputation problems in re-identifying human apparel by presenting a re-identification technique based on convolutional neural networks. Based on the inception structure, the framework provides a functions extractor and a more straightforward topology. High reputation scores are demonstrated by quantitative experiments, underscoring the need of specialized effort in re-identification.

Bedagkar-Gala, et.al,…[5] The study explores person re-identities across cameras in large-scale surveillance, focusing on spatiotemporal appearance models. Biometrics are employed for matching, however because of limitations in body charge or camera resolution, it is hard to recover face or gait. Individual re-identification involves dynamic probing sets and gallery updates, and it is an open set matching problem.

Y. Yan, et.al,…[6] This study proposes a lively pattern selection method using net pix for each image form, focusing on multimedia content evaluation. It proves that picture categories' active learning performance can be enhanced by textual content features. Traditional active sampling methods are limited to one kind of media; however, LUPI, a novel supervised mastery paradigm, can overcome this limitation. The work offers five ways to integrate uncertainty measurements and teaches SVMs about visible and weak features associated to text capabilities. By avoiding the exchange-off parameter between measurement approaches, the model is solved by the use of a unique optimization technique.

Y. Yang, et.Al,…[7] The goal of the study is to enhance feature choice performance through the utilization of data from related jobs. It emphasizes the value of multitask learning, adjusting inter-task knowledge, and exchanging records among tasks for supervised investigations. Although function selection improves performance, overall performance is not always improved. Specialized classifiers, such the linear Support Vector Machine (SVM), which assigns different weights to different characteristics, enhance feature selection procedures. The KTH dataset and KNN are used in the study as a stand-in classifier for motion recognition. After selecting features, KNN performs more accurately than SVM, presumably because SVM can weigh extraordinary qualities.

J. Luo, et.Al,…[9] The research suggests a framework for multimedia content analysis and retrieval that includes a brand-new transductive rating technique known as Ranking with Local Regression and Global Alignment (LRGA). Utilizing the sample distribution throughout the data set, LRGA learns a Laplacian matrix to rank the facts. In order to globally align local linear regression models from all information variables, it suggests a single objective feature. This enables each fact point to be given the best ranking score possible. Additionally, the paper suggests a semi-supervised reinforcement learning method that splits RF data into paired constraints and two enterprises. The algorithms demonstrate better performance than existing studies when tested in content-based full move-media retrieval and 3-dimensional motion/pose fact retrieval.

3. GRASSMANN MANIFOLDS AND SUBSPACE DISTANCE

To facilitate our explanation of Grassmann registration manifolds, we give a brief overview of canonical angles, subspace distances, and Grassmann manifold properties. There are thorough summaries of these subjects accessible[10].

3.1 Grassmann Manifolds

A Grassmann manifold An,q is a set of p-dimensional linear subspaces of Rn (q-planes in Rn) for 0 < q n. An,q = Un,q / Oq is the natural quotient representation of this Grassmann manifold, where Un,q is a Stiefel manifold (a set of n p orthonormal matrices) and Op is the orthogonal group. This representation states that two matrices belong to the same equivalence class if their columns span the same p-dimensional subspace. Thus, the entire equivalency class can be represented by the subspace spanned by the columns of a given matrix Y.

$$[\mathrm{Y}] = \{\mathrm{Y}Q_p : Q_p \in \mathrm{O}_p\} \qquad (1)$$

3.2 Canonical Angles

Given two vectors, m and n Rn, the angle between them is defined as

$$angle\{m,n\} = \arccos\frac{(m,n)}{||\mathrm{m}||||n||} \qquad (2)$$

Canonical angles are a set of angles between subspaces that can be defined recursively. Given two subspaces in Rn, [M] and [N], we may define the canonical angles [11], −k{[M],[N]}, recursively as

$$\cos(\Theta_k) = \max_{a\in[\mathbf{M}]} \max_{b\in[\mathbf{N}]} a^T b = a^T b_k \qquad (3)$$

subject to

$$||a|| = ||b|| = 1, a^T a_i = 0,, b^T b_i = 0, j = 1, \ldots, L-1$$

It is obvious that Θ is a vector of all canonical angles and ∠k{[M],[N]}∈[0, π /2].

3.3 Subspace Distance

The shortest path between two places in Grassmannian space is geodesic because the space is curved. The geodesic distance is defined using a collection of canonical angles, as demonstrated by Wong[12].

$$d_g(m,n) = ||\Theta||_2 \qquad (4)$$

But not everywhere is this distance differentiable. A chordal distance [13] is an alternate way to quantify geodesic distance.

$$d_g(m,n) = ||sin\Theta||_2 \qquad (5)$$

When the planes are close together, the chordal distance is differentiable everywhere and roughly corresponds to the geodesic distance. Our subspace distance metric is $d_c(m,n)$.

3.4 Formation of Registration Manifolds

Given that sampling and characterizing a registration manifold is the most crucial step in our recommended methodology. Given a pair of eye coordinates, we find a set of affine parameters for geometric normalization. The affine transformation maps the (m,n) coordinate of a source picture to the (a,b) coordinate of a normalized image. The transformation can be expressed as follows:

$$\begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} \cos(\Theta) & -\sin(\Theta) & dx \\ \sin(\Theta) & \cos(\Theta) & dy \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1+\mathrm{tx} & 0 & 0 \\ 0 & 1+ty & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} m \\ n \\ o \end{pmatrix} ..(6)$$

When the third matrix indicates scaling, the second matrix shows skew, and the first matrix describes translation and rotation. More compactly, these converted coordinates can be rewritten as:

$$\begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} q1 & q3 & q5 \\ q2 & q4 & q6 \end{pmatrix} \begin{pmatrix} m \\ n \\ 1 \end{pmatrix} \qquad (7)$$

The affine transformation has six control parameters, as shown by equation (7). Equation (8) illustrates how a series of registration images is sampled in this research by varying these six affine parameters.

$$\begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} q1 + \Delta q1 & q3 + \Delta q3 & q5 + q5 \\ q2 + \Delta q2 & q4 + \Delta q4 & q6 + \Delta q6 \end{pmatrix} \begin{pmatrix} m \\ n \\ 1 \end{pmatrix} \qquad (8)$$

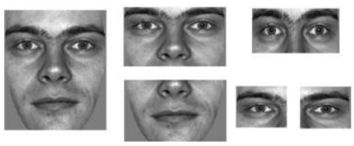
Despite the fact that local features can be divided into local regions and local elements [14], the full face is considered in a holistic image. Typically, local components consist of an eye or another distinguishing facial feature. Regular sampling procedures, including laying a grid over the face, are used to produce local areas. In this work, we employ a holistic image, local regions, and local components for face identification. Our choices for the local components are the eyes, the top and lower faces, and the left and right eyes. Local regions are the 3 × 3 and 5 × 5 face windows.

Fig 1: An illustration of a local component, a holistic face, and local areas

4. FACE RECOGNITION SYSTEM USING GRASSMANN ALGORITHM

The task of distinguishing individual faces and associating them with recognised identities falls to this module. Every face is given a unique digital signature created by machine learning algorithms that may be cross-referenced with a database of recognised faces. Although Grassmann learning for face recognition is a computationally demanding method, it has been demonstrated to work well in a number of situations. In order to ensure accurate feature extraction, it is crucial to make sure the preprocessing methods are consistent and the dataset is large enough to develop a trustworthy model. Using a web camera, the student's current facial image is collected in real time, and feature values are extracted. After that, use the classification procedure to determine whether or not the student's facial image was present.

A common mathematical method for analysing and comparing facial characteristics in face recognition is the Grassmann algorithm. The following are the fundamental procedures for using the Grassmann algorithm to facial recognition:



4.1 Feature extraction:

First, facial traits must be extracted from the photos of the faces that need to be recognised.

4.2 Face representation:

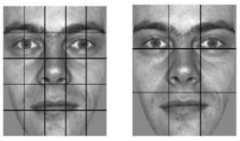
After that, a high-dimensional space is used to represent the retrieved face characteristics as points. This area is frequently called a face space or a feature space.

4.3 Grassmann manifold:

The Grassmann manifold is a mathematical representation of each viable subspace of a given dimension in a high-dimensional space. The Grassmann technique compares the subspaces that represent the different faces' facial features on this manifold.

4.4 Subspace projection:

Each face is represented as a subspace in the feature space. Then, using the Grassmann method, these subspaces are projected onto the Grassmann



manifold, producing a set of comparable points.

4.5 Distance computation:

Finally, the distance between the subspaces is computed using a distance metric such as the Grassmann distance. The distance metric takes the Grassmann manifold's geometry into consideration when determining how similar or unlike the subspaces are from one another.

4.6 Classification:

The faces are then categorised as belonging to known or unknown people using the calculated distances. In this stage, a distance metric threshold value is usually defined, above which a face is deemed unknown.
All things considered, the Grassmann algorithm is an effective face recognition method that can adapt to changes in illumination, posture, and facial emotions. Because it can accurately represent the diversity of face characteristics in a low-dimensional space, it is especially helpful when there are few training examples. The suggested system architecture diagram is shown in Fig.2.

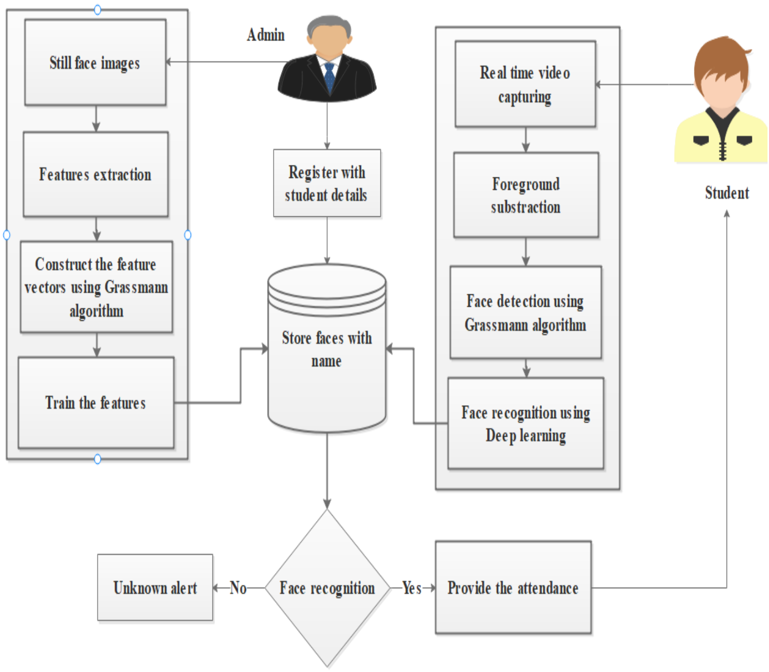


Fig 2: System Architecture for attendance management system

The proposed modules are shown below

I. FRAMEWORK CONSTRUCTION

This project uses a methodical foundation. Establish project goals and assemble a knowledgeable team to get the project started. Gather and annotate a variety of datasets while using picture improvement pre-processing techniques. Align faces that are found by using facial detection methods. Utilize a Convolutional Neural Network (CNN) that has been trained beforehand to extract features. Use the Grassman still-to-video matching algorithm and modify it so that it can compare feature vectors. Create a database with well-known features so that learning can continue. Establish thresholding for identification, integrate communication modules for timely notifications, and implement a real-time processing pipeline. Provide an intuitive user interface to facilitate efficient system monitoring. For continued effectiveness, implement the system, carry out extensive testing, and create maintenance procedures.

II. FEATURE VECTOR CONSTRUCTION

Facial feature extraction in this project requires a few crucial processes. Prior to picture augmentation, facial regions are first identified and pre-processed. After the faces are aligned, they are sent into a Convolutional Neural Network (CNN) that has already been trained. This CNN is denoted as $F=CNN(I;\theta)$, where I is the input picture, θ is the CNN parameters, and F is the feature vector. Using the Grassmann distance calculation, the Grossmann algorithm can further match features between still photos and video frames. Feature vectors are compared with a similarity metric against a database (Y). Real-time attendance management is made easier by the identification result being calculated based on a predetermined threshold.

III. FACIAL ENTROLLMENT

This process involves capturing and registering individuals' facial data to establish a comprehensive database. During facial enrollment, the system acquires facial images or video frames, extracting pertinent features crucial for subsequent identification. The term 'registration' signifies the formal inclusion of these facial features into the system's database, often associating them with unique identifiers. This integral step initializes the face recognition system, laying the groundwork for accurate identification and verification of individuals based on their registered facial characteristics. The process underscores the systematic on boarding of facial data essential for the system's functionality.

IV. CLASSIFICATION

The described face recognition-based attendance system classifies as an innovative application of advanced biometric technology. This real-time, intelligent solution utilizes webcams to capture still images and video recordings, providing a seamless method for generating accurate attendance records. The classification is marked by its departure from traditional methods, leveraging facial identification to notify parents and administrative teams promptly. Notably, the incorporation of the Grassmann algorithm for still-to-video matching stands as a pivotal element, employing deep learning to enhance accuracy in real-time identification of unknown matches. The project thus falls into a transformative category, revolutionizing attendance management through cutting-edge technology and efficient communication mechanisms.

V. ATTENDANCE SYSTEM

Utilizing webcams, it captures still images and video recordings, employing facial identification to seamlessly generate precise attendance records. This module diverges from traditional methods by promptly notifying parents of children's attendance status through messages, enhancing communication. The incorporation of the Grassman algorithm for still-to-video matching ensures deep learning-based accuracy in identifying unknown individuals. This module acts as a transformative element in attendance tracking, revolutionizing administrative efficiency by delivering timely results. Its intelligent design addresses the challenges faced by institutions, fostering an advanced and communicative attendance management system.

5. CONCLUSION

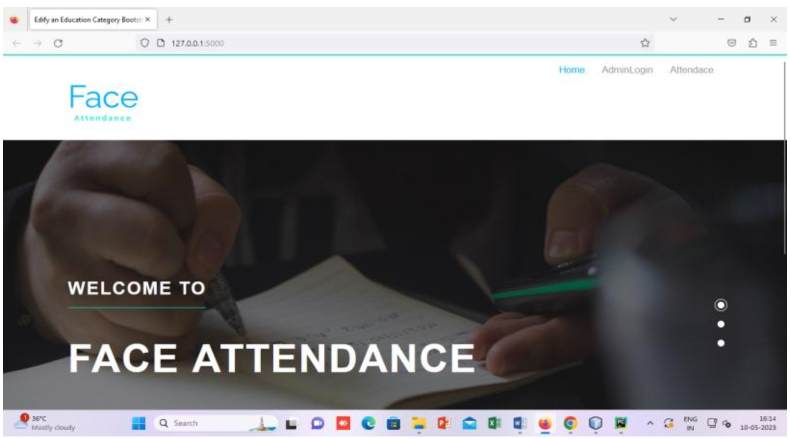
This research examined facial recognition methods for both video and still photos. The majority of these current methods only do still picture face recognition or video-to-video matching and require well-aligned face images. They are inappropriate for facial recognition in surveillance scenarios for the following reasons: restriction on the quantity of faces shots (about 10) that might be taken from each film because of the drastic changes in lighting and posture; due to low video quality and resource limitations brought on by real-time processing, there was no guarantee of facial picture alignment. In order to do face identification in still images and videos while under surveillance, a local facial feature-based framework might be proposed. This generic framework can do real-time face-to-face matching. While training uses static photos, the recognition challenge is conducted over video sequences. Our results based on the Grassmann and Convolutional Neural Network technique showed higher recognition rates when video sequences were used instead of statics. This technique is tested with real-time picture datasets and an SMS alert system for face detection in still photos and videos.

6.EXPERIMENTAL RESULTS

In this simulation, provide real time face datasets in student attendance system using Python framework as front end and MySQL as back end.

Fig 3: Home page

In this screen show the home page for student attendance system



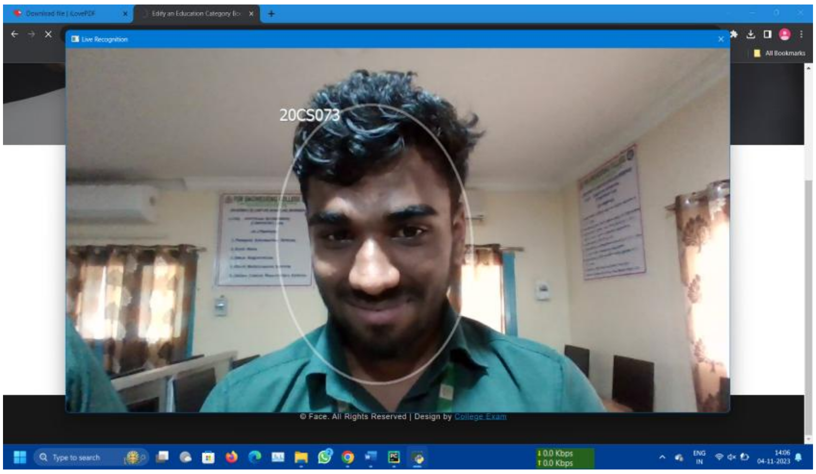
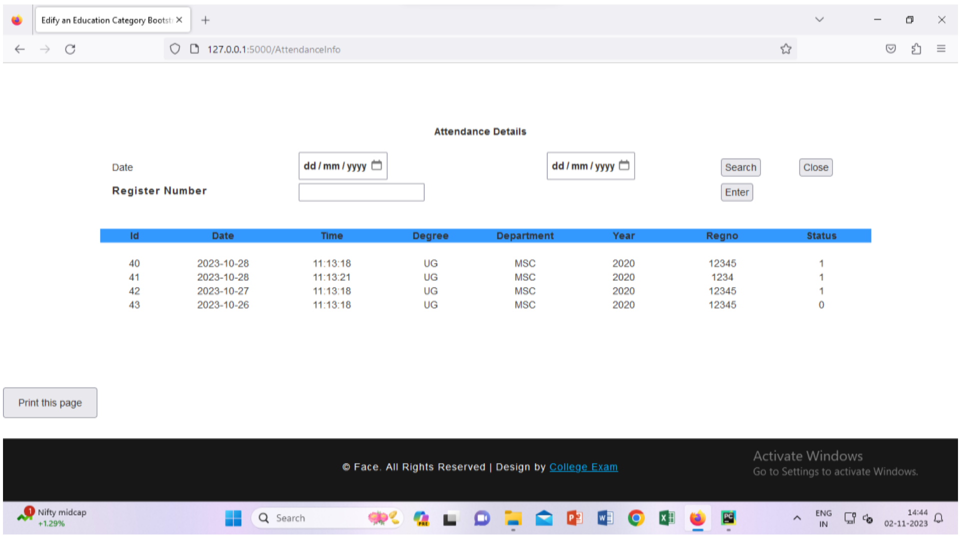


Fig 4: Attendance for real time face

Real time face datasets are used to train the person details in terms of feature vectors



Id	Date	Time	Degree	Department	Year	Regno	Status
40	2023-10-28	11:13:18	UG	MSC	2020	12345	1
41	2023-10-28	11:13:21	UG	MSC	2020	1234	1
42	2023-10-27	11:13:18	UG	MSC	2020	12345	1
43	2023-10-26	11:13:18	UG	MSC	2020	12345	0

Fig 5: Storing student records in database

Facial features are stored with labels in the data base

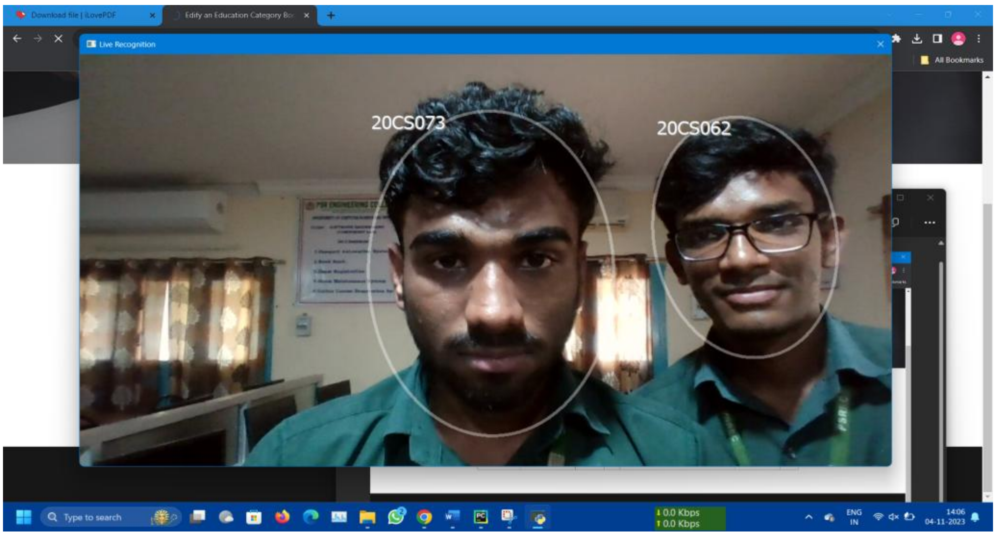


Fig 6: Multi person attendance

Achieve multiple person face attendance in single frame.

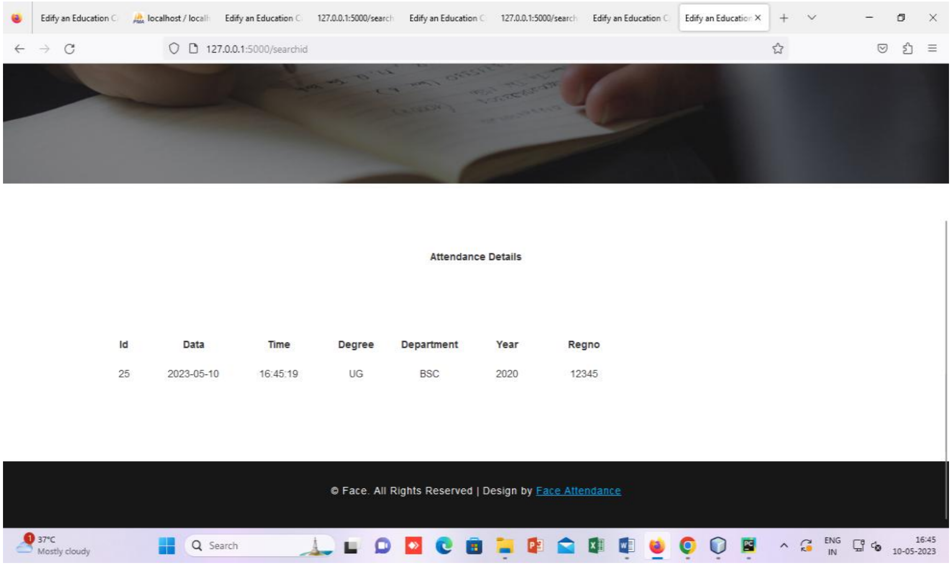


Fig 7: Attendance report

This shows the overall attendance report for each and every students

Fig 8

8

: Alert System

Notify the parents via SMS

7.REFERENCES :

[1] M. Ayazoglu, B. Li, C. Dicle, M. Sznaier, and O. Camps. Dynamic subspace-based coordinated multicamera tracking. In 2011 IEEE International Conference on Computer Vision(ICCV), pages 2462–2469, Nov. 2011.

[2] D. Baltieri, R. Vezzani, and R. Cucchiara. Learning articulated body models for people re-identification. In Proceedingsof the 21st ACM International Conference on Multimedia, MM ’13, pages 557–560, New York, NY, USA, 2013. ACM.

[3] D. Baltieri, R. Vezzani, and R. Cucchiara. Mapping appearance descriptors on 3d body models for people reidentification. International Journal of Computer Vision, 111(3):345–364, 2015.

[4] I. B. Barbosa, M. Cristani, B. Caputo, A. Rognhaugen, and T. Theoharis. Looking beyond appearances: Synthetic training data for deep cnns in re-identification. arXiv preprintarXiv:1701.03153, 2017.

[5] A. Bedagkar-Gala and S. Shah. Multiple person reidentification using part based spatio-temporal color appearance model. In Computer Vision Workshops (ICCVWorkshops), 2011 IEEE International Conference on, pages 1721–1728, Nov 2011.

[6] 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

[7] 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.

[8] 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.

[9] 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.

[10] 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.

[11] A. Bedagkar-Gala and S. K. Shah. Part-based spatiotemporal model for multi-person re-identification. PatternRecognition Letters, 33(14):1908 – 1915, 2012. Novel Pattern Recognition-Based Methods for Re-identification in Biometric Context.

[12] J. Berclaz, F. Fleuret, E. T¨uretken, and P. Fua. Multiple object tracking using k-shortest paths optimization. IEEETransactions on Pattern Analysis and Machine Intelligence, 2011.

[13] K. Bernardin and R. Stiefelhagen. Evaluating multiple object tracking performance: the CLEAR MOT metrics. EURASIPJournal on Image and Video Processing, (246309):1–10, 2008.

[14] L. Beyer, S. Breuers, V. Kurin, and B. Leibe. Towards a principled integration of multi-camera re-identification and tracking through optimal bayes filters. CVPRWS, 2017.

[15] M. Bredereck, X. Jiang, M. Korner, and J. Denzler. Data association for multi-object Tracking-by-Detection in multicamera networks. In 2012 Sixth International Conference onDistributed Smart Cameras (ICDSC), pages 1–6, Oct. 2012.