Smart Attendance Monitoring System (SAMS): A Face Recognition based Attendance System for Classroom Environment

Shubhobrata Bhattacharya, Gowtham Sandeep Nainala, Prosenjit Das and Aurobinda Routray

***Abstract*— In present academic system, regular class attendance of students' plays a significant role in performance assessment and quality monitoring. The conventional methods practised in most of the institutions are by calling names or signing on papers, which is highly time-consuming and insecure. This article presents the automatic attendance management system for convenience or data reliability. The system is developed by the integration of ubiquitous components to make a portable device for managing the students' attendance using Face Recognition technology.**

***Index Terms*— Attendance Monitoring, Face Quality Assessment, Face Recognition.**

I. Introduction

Attendance plays a pivotal role in determining academic performance of children and youth in schools and colleges. The regularity of attendance shows that the students are less likely to engage in delinquent or destructive behaviour [1], [2]. Chronic absence increases the risk of school failure and early dropout. [3], [4]. Manual maintenance of attendance are inefficient due to the following reasons:

- It takes away a lot of lecture hours
- Prone to proxies or impersonations

To resolve this problem of attendance, many attendance management systems have been introduced in recent years. Jain *et. al* [5] developed a desktop based application in which students are given attendance by clicking a checkbox next to their name and then by clicking the register button to mark their presence. In 2013, Bhalla *et. al* [6] have proposed blue-tooth based attendance system. Application software installed in mobile phone enables to register the attendance via blue-tooth connection and transfer the notification to the instructor. Works of [7] propose a system for employee attendance based on the fingerprint. The system compares one fingerprint template with all previously stored in the database. In [8], Joardar *et. al* has developed an attendance system based on the palm dorsal subcutaneous vein pattern of individuals.

However, most of these systems have respective limitations in portability, accessibility, authenticity or cost. So an endeavour to overcome the shortcomings of the respective systems leads to the development of a Smart Attendance Monitoring System (SAMS) based on face recognition. Unlike other biometric and non-biometric means of attendance system, face recognition technology stands tall with its unique advantages. Every student has a separate facial identity and it can not be faked by mere proxies. Moreover, the class teachers feel more acquainted with the student by there countenance than the name or roll number. These works have the following contributions:

- Best face selection method using face quality assessment and robust face representation using deep convolution network.
- Portable device based on embedded systems for attendance in classrooms.

The paper is organized as follows. Section II presents the proposed framework. Several steps of the framework are explained in details within this section. Evaluation of the work is provided in Section III. Section IV concludes the paper.

II. Methodology

Having a video sequence as the input to the system the details of face detection, facial features extraction, normalization of facial features and quality score assignment are described in the following subsections.

A. Face Detection

For better accuracy of face-log generation, we employed face tracking technique. All we did was first detect the face using Viola & Jones idea as described in [9] and then, we used the correlation tracker from the dlib library to keep track of the face from frame to frame. This approach also saves computational power since we don't have to detect the face after transforming to a new frame in the real-time video sequence. This helps to generate a face-log i.e, a concise representation of the face of the subject in a video sequence [10].

B. Parameters

1) Pose estimation: Since people move around and look at different directions in front of the real-time camera, it is possible to have a wide range of head poses oriented at different angles. But for the sake of biometrics, it is important to have least rotated face as a standout in the entire face-log. Thus it is important to include this feature in face quality assessment. We determined the head pose using three angles: Roll, Yaw, Pitch. All these angles are typically between -90 to + 90. The roll and pitch are adjusted by aligning technique during face-log generation, so our only concern is yaw angle. Using face landmarks detection, we calculated the coordinates of nose tip and also the point between the eyebrows. If (x_1, y_1) and (x_2, y_2) are such points, then yaw angle is computed as:

$$yaw = abs(\arctan 2(y_2 - y_1, x_2 - x_1)) \tag{1}$$

2161-377X/18/$31.00 ©2018 IEEE
DOI 10.1109/ICALT.2018.00090

TABLE I: Head-pose estimation

Range Of Yaw Angle	0 to 10	10 to 20	20 to 30	More than 30
NHP	1	0.5	0.33	0.25

Then Normalized head-pose parameter (NHP) is obtained as shown in table,

2) Sharpness: It is very likely to have blurry images in real time video sequences because the faces are moving. Thus it is important to include this feature in face quality assessment. To compute the sharpness of an image, we utilized the variance of an image Laplacian. This can be defined as:

$$Sharpness = \sum_{(i,j)\in\mho(x,y)} (\triangle I(i,j) - \overline{\triangle I})^2 \qquad (2)$$

where $\overline{\triangle I}$ is the mean value of image Laplacian within $\mho(x, y)$. We took a grey-scale channel of an image and convolved it with the $(3x3)$ kernel and took the variance of that result.

$$kernel = \begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix} \qquad (3)$$

If the variance falls below a pre-defined threshold, then the image is considered as a blur. This threshold depends on the working environment and can be set accordingly. We then normalized it with a pre-defined threshold,

$$NormalizedSharpness(NS) = \frac{Sharpness}{threshold} \qquad (4)$$

3) Image size or resolution: Since we employed the face tracking technique, the camera tracks the face as long as the face is in the scene. But, as the face moves far away from the camera, there will be a large distance between the camera and the face. Thus, the size of the face becomes smaller. Thus, it is important to include this feature in face quality assessment. Using face landmark detection, we calculated the position of the eye corners in a face. Let (x_L, y_L) be the coordinates of the left eye corner and (x_R, y_R) be the coordinates of the right eye corner. The distance between them is given by:

$$Resolution = \sqrt{(x_L - x_R)^2 + (y_L - y_R)^2} \qquad (5)$$

Then we normalized the obtained resolution with the threshold for interocular distance as,

$$NormalizedResolution(NR) = \frac{Resolution}{threshold} \qquad (6)$$

Larger the distance, smaller will be the size of the face and hence, lesser will be the resolution.

TABLE II: Weights for various parameters in computing FQA

Parameter	Head-pose	Sharpness	Brightness	Resolution
Weight	17	9	6	8

4) Brightness: Changes in Lightening conditions are quite common in real-time applications like surveillance cameras. It is easier to apply Local feature extraction on brighter faces than on darker faces. Thus, it is necessary to include this feature in face quality assessment. To obtain this parameter, we calculated the mean of all the intensities of various channels (R, G, B) present in the image.

$$Brightness = (R + G + B)/3 \qquad (7)$$

$$NormalizedBrightness(NB) = Brightness/100 \qquad (8)$$

C. Final Score

In order to get the best quality image in the real-time video sequence, we need to assign weights to each of the normalized parameters (NHP, NS, NR, NB). We gave the highest priority to the head pose followed by other parameters as shown in Table II. So, we computed the Face Quality Assessment (FQA) as,

$$FQA = NHP \times 17 + NS \times 9 + NB \times 6 + NR \times 8 \qquad (9)$$

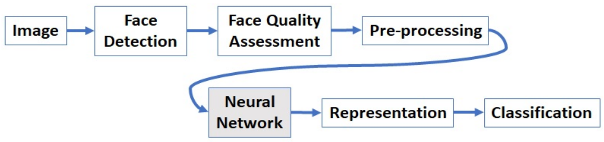
Greater is the value of FQA, greater will be the quality of face which will be stored in Face-log. We can set the threshold for these FQA values of generated images to prevent the bad quality images from being stored in face-log.

D. Representation : Deep Learning

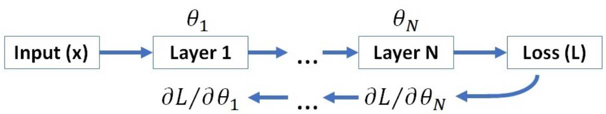
Face representation is the core of the recognition algorithm used in this system. The face image captured after the quality assessment is needed to be represented in form of feature for further processing. The preprocessed images are too high-dimensional for a classifier to take directly on input. To obtain a low-dimensional distinct feature from the face images we used Convolution Neural Network (CNN), popularly known as deep learning. A deep network is a feed-forward network comprising of many function compositions, or layers. The network is provided with a loss function $\mathcal{L}$. The loss function measures how accurately the neural network classifies an image. Common layer operations repeated on the input image over the sequence of layers are:

- Spatial convolutions that slide a kernel over the input feature maps.
- Linear or fully connected layers that take a weighted sum of all the input units, and
- Pooling that takes the max, average, or Euclidean norm over spatial regions.

These operations are often followed by a nonlinear activation function, such as Rectified Linear Units (ReLUs) [11], which are defined by $f(x) = max\{0, x\}$. Neural



(a)



(b)

Fig. 1: (a) Pipe line of the system; (b) Block diagram of the deep neural network

network training is an optimization problem that finds a θ that minimizes (or maximizes) $\mathcal{L}$. With differentiable layers, $\mathcal{L} = \theta_i$ can be computed with back-propagation. The optimization problem is then solved with a first-order method [12], which iteratively progress towards the optimal value based on $\mathcal{L} = \theta_i$. Figure 1b shows the logic flow for face recognition with neural networks. The neural network predicts some probability distribution and the loss function $\mathcal{L}$ measures how well is the prediction close to the person's actual identity.

E. Hardware Requirements

Hardware components used for the making of SAMS are as follows.

- UDOO X86 Ultra Single Board Computer used for running the algorithms.
- 15MP Camera.
- Mini-keyboard.
- 7-inch LCD.
- Portable power bank and $5V$ to $12V$ DC converters

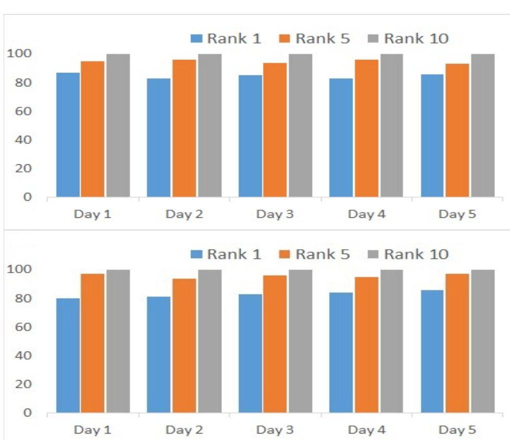


Fig. 2: Performance of SAMS for Classroom 1 and Classroom 2

III. Performance of the System

SAMS has been designed to register the face of each individual for the first time. Once done, the network trains it automatically for future usage. For the next classes, the students can get their self-attendance done with the GUI offering a drop-down menu for the recognized face. This is because of the chance of look alike within the class. The first name in the drop down has the highest probability of the match. We stated this as Rank 1 accuracy. For the closed match within 5 faces, we call it Rank 5. The performance of the system is shown in Fig. 2.

IV. Conclusion

An automatic attendance management system aims at solving the issues of manual methods of existing systems. We have used the concept of face recognition to implement a system that marks the attendance of a particular person by detecting and recognizing the face. These systems perform satisfactorily with different facial expressions, lighting and pose of the person. There is room for improvement since these systems sometimes fail to recognize every face student present in the classroom. We have made the device portable for easy use even when the sessions are on, without disturbing the class. There are future scopes to make a more compact ergonomics to make it a more user-friendly product to make an impact in building a more healthier academic environment.

References

[1] J. L. Epstein and S. B. Sheldon, "Present and accounted for: Improving student attendance through family and community involvement," *The Journal of Educational Research*, vol. 95, no. 5, pp. 308–318, 2002.

[2] D. D. Ready, "Socioeconomic disadvantage, school attendance, and early cognitive development: The differential effects of school exposure," *Sociology of Education*, vol. 83, no. 4, pp. 271–286, 2010.

[3] C. Bruner, A. Discher, and H. Chang, "Chronic elementary absenteeism: A problem hidden in plain sight," *Attendance Works and Child & Family Policy Center*, 2011.

[4] C. P. McCluskey, T. S. Bynum, and J. W. Patchin, "Reducing chronic absenteeism: An assessment of an early truancy initiative," *NCCD news*, vol. 50, no. 2, pp. 214–234, 2004.

[5] S. K. Jain, U. Joshi, and B. K. Sharma, "Attendance management system," *Masters Project Report, Rajasthan Technical University, Kota*, 2011.

[6] V. Bhalla, T. Singla, A. Gahlot, and V. Gupta, "Bluetooth based attendance management system," *International Journal of Innovations in Engineering and Technology (IJIET) Vol*, vol. 3, no. 1, pp. 227–233, 2013.

[7] S. S. Mahat and S. Mundhe, "Proposed framework: College attendance management system with mobile phone detector," *International Journal of Research in IT and Management*, vol. 5, no. 11, pp. 72–82, 2015.

[8] S. Joardar, A. Chatterjee, and A. Rakshit, "A real-time palm dorsa subcutaneous vein pattern recognition system using collaborative representation-based classification," *IEEE Transactions on Instrumentation and Measurement*, vol. 64, no. 4, pp. 959–966, 2015.

[9] P. Viola and M. J. Jones, "Robust real-time face detection," *International journal of computer vision*, vol. 57, no. 2, pp. 137–154, 2004.

[10] K. Nasrollahi and T. B. Moeslund, "Complete face logs for video sequences using face quality measures," *IET signal processing*, vol. 3, no. 4, pp. 289–300, 2009.

[11] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[12] Y. Bengio, I. J. Goodfellow, and A. Courville, "Deep learning," *Nature*, vol. 521, pp. 436–444, 2015.