Real-time Monitoring of Workforce: An approach based on Deep Features

Sadique K.M.
Department of MCA Maharaja Institute of Technology Mysore, Mandya, Karnataka-India

Amos R.
Department of MCA Maharaja Institute of Technology Mysore, Mandya, Karnataka-India

ABSTRACT
In this paper, we monitor real-time workforce attendance. At first, we record the check-in and check-out of the workforce. Next, keep track of their movements at various premises within the organization. Finally alarm the administrator for unauthorized movement. In order to meet these requirements, we extracted state-of-the-art deep learning-based features by utilizing AlexNet. Extensive experiments were conducted on our created dataset. From the experiments it was revealed that extracted features substantially perform better.

Keywords
Real-time monitoring, attendance system, unauthorized movement, deep learning, AlexNet.

1. INTRODUCTION
Ever since 9-11 terrorist attack governments/organizations became serious about security of their organizations, homes, and business setups, etc. In view of this, different measures have been taken into account, face recognition is one such approach. There exist numerous works towards authorizing the work-force based on face recognition in an organization. That varies from traditional hand-crafted features to state-of-the-art deep learning-based features [1, 2]. Although, significant accuracy has been achieved, but still there is a plenty of scope to further enhance/validate the existing system. Or, test their precision in different scenarios and setups. With this motivation, we explore the state-of-the-art deep learning-based features for our organization to authorize and monitor work-force.

The existing systems are mainly traditional/manual based systems. Wherein different log-files are maintained to record check-in and check-out of work force. With the passage of time, several automatic thumb-impression based systems came into existence [3], however, they are capable only to record check-in and check-out. So, need was felt for complete automatic monitoring system. That would be capable of carrying out all necessary tasks. For example, check-in and check-out, deviation from authorized routes, etc. In view of this, in this paper, our aim is to design/develop a real-time attendance monitoring system which would be capable of carrying out stated functionalities/objectives.

The rest of the paper is organized into three sections. In section 2 detailed description is given for feature extraction. Section 3 gives extensive experiments and finally, conclusion and future work is drawn in section 4.

2. Proposes System
The pipeline of the proposed system is shown in Fig.1. The main sections of the proposed system are defined as follows.

2.1 Pre-processing
There are various pre-processing steps that precede feature extraction step. These are listed in sequence.

2.1.1. Key-frame extraction
In order to create the video summary or compress the video content we extracted key-frames with the help of k- means clustering algorithm. We kept the value of $k = 3$ for coherence and simplicity. Finally, from each cluster a representative frame is chosen/selected as a key-frame.

2.1.2. Face Detection
This step is crucial for the efficacy of the proposed system. That is why we have adopted Viola-Jones face detection algorithm [4]. The motivation for that is two-fold: first it works well in real time scenarios, second, accuracy with which it detects the faces.

2.1.3. Filtering, smoothening and resizing
To remove the noise from the detected faces (step 2.1.2), we have used median filtering technique [5], with a 3×3 window. A non-linear filter which has proven its robustness from various studies. For smoothening purpose, we have used Gaussian filtering technique [5]. Since, we are using AlexNet for extraction of deep features that expects $224 \times 224 \times 3$ image size. Therefore, we resized filtered, and smoothened face images to $224 \times 224 \times 3$ dimensions.

2.2. Feature Extraction
To extract deep learning-based features, we have used state-of-the-art AlexNet [6], convolutional neural network architecture. Whose efficacy has been demonstrated from

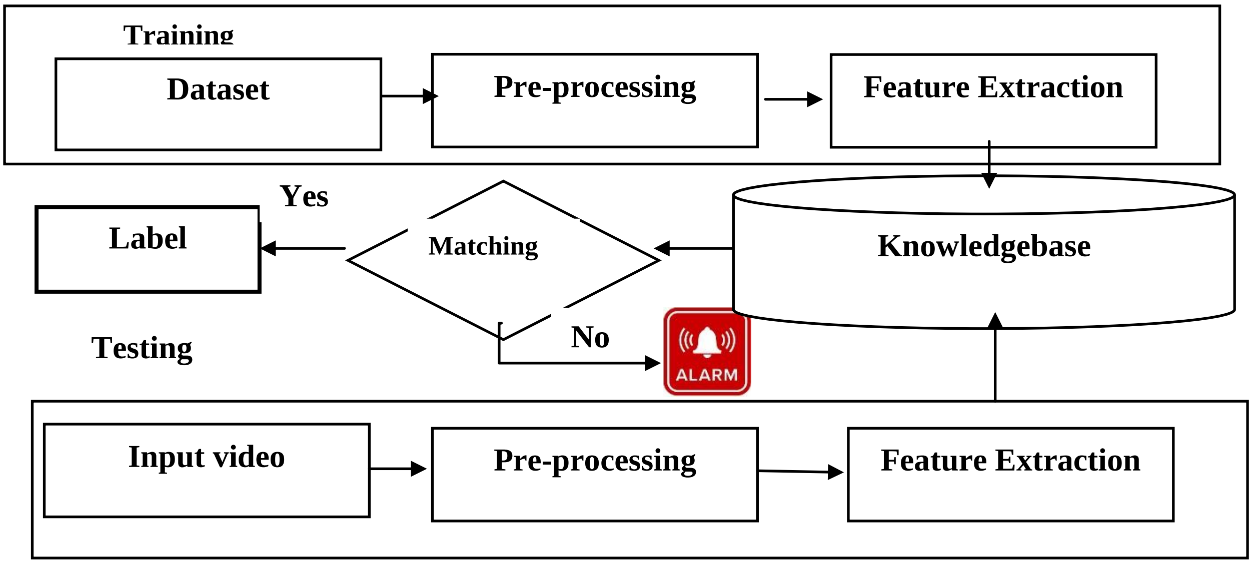

Fig.1: Illustration of the proposed system

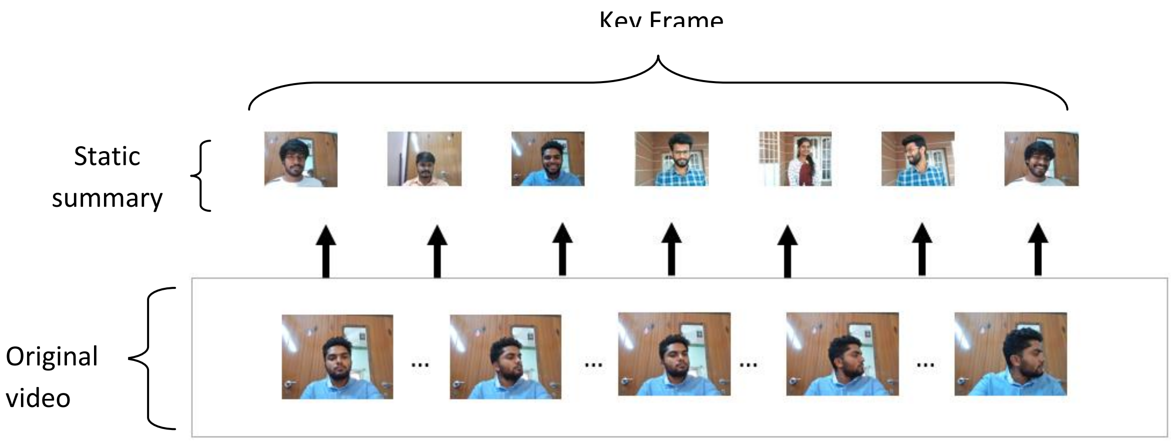

Fig.2: Illustration for Key frame extraction

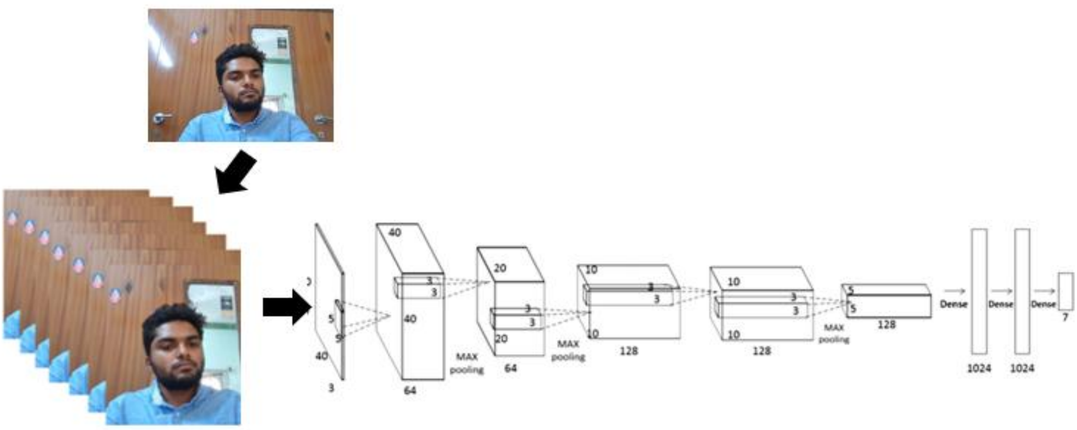

Fig. 3: Illustration of extraction of deep features from AlexNet [6]

various problem domains. It consists of eight layers (8) of which five (5) are convolutional layers, and three (3) are fully connected layers, as shown in Fig.2. The internal representation at the convolutional layer two is shown in Fig. 3.

2. EXPERIMENTATIONS
2.1. Dataset

For this study we have created our own dataset. At outset we have taken three-minute (3) videos from the camera installed at various locations in our organization (Fig. 4). We have divided our dataset in five (5) classes each with 150 instances/samples. Afterwards, AlexNet is used to extract the features (see Section 2). Note, we have tuned the AlexNet to our created dataset. Any outside person will be labelled as intruder or alert will be sounded to an administrator. Fig. 4 shows the sample of our created dataset.

2.2. Experimental results

To corroborate the efficacy of the proposed system we have used accuracy measure as an evaluation measure, shown in Equation 1.

$$\text{Accuracy}=\frac{TP+TN}{TP+TN+FP+FN} \qquad \textbf{(1)}$$

Fig.4: Sample shot of the created dataset

Table 1. Average class wise recognition results shown in percentage

Class Index	Training: Testing							Class wise average result	
	Trail 1		Trail 2		…	Trial 20			
	70:30	80:20	70:30	80:20	…	70:30	80:20	70:30	80:20
1	91.78	93.58	91.75	93.98	…	91.89	94.50	**91.80**±0.01	94.02 ±.01
2	90.69	91.85	89.96	91.59	…	89.99	92.38	92.54 ±.01	93.94 ±.01
3	92.58	93.89	91.99	93.98	…	92.98	93.99	92.51 ±.10	93.95 ±.01
4	92.48	93.69	92.98	93.78	…	91.98	93.59	92.48 ±.01	93.68 ±.10
5	93.58	94.58	92.96	94.89	…	92.99	94.98	92.17 ±.10	94.81 ±.11
Overall Average result								**91.70 ±.011**	**94.68 ±.11**

Where, TP refer true positive i.e., label is positive and prediction is also positive, TN true negative i.e., label is negative and the prediction is also negative, FP false positive i.e., label is negative but the prediction is positive, and FN as false negative i.e., label is positive but the prediction is negative. Note we have divided the dataset tuning and testing in the ratios of and 70:30 and 80:20, respectively. We repeated experiments for 20 random trails and recorded accuracy and average accuracy is computed from 20 trails for the corroboration of the results. Table 1 shows the results. If the entrant is intruder, then he/she will not be recognized/classified simultaneously an alarm is sent to the administrator. From the experiments we were able to alarm the administrator at each misclassification.

We observe from results class Index 2 is showing less result as compared to other classes. This might be due to the mask used by the person. We also observe that, as we increase the samples for training, we also observe the increase in recognition process. For our created dataset, we observe the recognition results from 91% to 94%. That is also observed from number of other applications, like in [7-8]. Note, the person/class which is not allowed to walk on a particular route, is treated simply as intruder on that route. On the other route if he is allowed to walk on that route, he is a valid class. Similar procedure is repeated for other classes and routes

3. CONCLUSION AND FUTURE WORK

In this paper our aim was to build real time system for monitoring the workforce in an organization. In which, we controlled entry-exit of workforce. Further any un-authorized routes taken by a work force is also monitored. For this, we explored the AlexNet for feature extraction. Which extracts the deep learning-based features. From the experimentation it has been shown that proposed approach performs well on our created dataset. In future, we will test our proposed model by enhancing our dataset with large samples and classes. Next, we will also test our model with other Neural network-based architectures like Generative adverse neural networks (GANs), encoder-decoder based neural networks, in order to, test the robustness of proposed model.

4. REFERENCES

[1] Kavita and Manjeet Kaur, A survey paper for face recognition. International Journal of Scientific and Research Publications, Volume 6, Issue 7, July 2016.

[2] Baback Moghaddam. Principal Manifolds and Probabilistic Subspaces for Visual Recognition. IEEE Transactions on pattern analysis and machine intelligence, vol 24, no 6, 2002.

[3] S. S. Thokal1, Aarti N. Bhosale et al. Smart Ration Card with RFID, Biometrics and Sensors. International Journal of Computer Sciences and Engineering. Vol 6, 5, 2019.

[4] Shaily Pandey, and Sandeep Sharma. An Optimistic Approach for Implementing Viola Jones Face Detection Algorithm in Database System and in Real Time. International Journal of Engineering Research & Technology (IJERT) . Vol. 4 Issue 07, July-2015

[5] Rafael Gonzalez and Richard Woods. Digital Image Processing. Pearson, print, 2017

[6] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. Image Net Classification with Deep Convolutional Neural Networks. Communications of the ACM Volume: 60, Issue: 6, pp 84-90

[7] Mohammed Idrees Bhat, Sharada B, Sk. Md Obaidullah and Mohammed Imran.17th International Conference on frontiers in Handwriting Recognition (ICFR) PP-234-239-2020.

[8] Mohammad Idrees Bhat and B. Sharada. Automatic Recognition of Legal Amounts on Indian Bank Cheques: A Fusion-Based Approach at Feature and Decision Levels. International Journal of Computer Vision and Image Processing (IJCVIP) Vol. 10 Issue 4 pp-57-73.

5. AUTHRO'S PROFILE

Sadique K M pursuing his Masters in Computer Applications at Maharaja Institute of Technology Mysore, His research interests include computer vision and machine learning.

Amos R, Assistant Professor, Department of Masters in Computer Applications at Maharaja Institute of Technology Mysore. His research Interests include Computer vision, machine learning, and data science.