Pimpri Chinchwad Education Trust's
Pimpri Chinchwad College of Engineering (PCCoE)
(An Autonomous Institute)
Affiliated to Savitribai Phule Pune University(SPPU)
ISO 21001:2018 Certified by TUV

FA-2 Report

· **Name:** Pranav Shah
· **Branch:** Information Technology
· **Division:** B
· **PRN No.:** 124B2F016
· **Course Name :** Fundamentals of Digital Image Processing

Contexa Vision – AI-Powered Smart Workplace Monitoring System

1. Problem Statement

Conventional CCTV systems only record video footage and rely on human operators to interpret events, which is time-consuming, inconsistent, and often ineffective in real time. Traditional motion sensors and alarms frequently cause false triggers and fail to distinguish With human and animals

Furthermore, existing systems cannot assess workplace productivity or detect employee activity states (e.g., working, idle, away), which are critical for modern smart workplaces.

Therefore, a unified intelligent monitoring system is needed — one that can:
- Detect, recognize, and track humans in real time for security and attendance purposes.
- Monitor workplace activity and send alerts when unauthorized intrusions or employee idleness.
- Reduce false alarms and improve efficiency using AI-driven video analytics.

2. Motivation

1. Conventional CCTV systems are limited to continuous recording and require dedicated staff for monitoring, often leading to fatigue and missed incidents.
2. Modern workplaces demand automation, accuracy, and intelligent decision-making without human supervision.
3. Integrating AI-based surveillance with workplace analytics helps improve both security and productivity
4. A unified platform for attendance, activity recognition, and intruder detection ensures higher safety, transparency, and performance.

3. Objectives

1. Intruder detection using object and human recognition.
2. Face recognition for employee identification and attendance.
3. Activity recognition to monitor employee engagement (working, idle, away).
4. To generate real-time alerts and notifications to managers or authorities upon detection of unusual events or inactivity.
5. To maintain a secure and privacy-aware database of authorized personnel and activity logs.
6. To provide data-driven insights for workplace efficiency, safety, and behavior analytics.
7. To minimize false alarms using advanced AI and computer vision techniques

4. Introduction

Security and productivity are two key challenges in modern organizations. Traditional CCTV setups merely record video footage without analysis, relying entirely on human operators to identify potential issues. This not only increases operational costs but also introduces human error and delayed response times.

Contexa Vision aims to bridge this gap by implementing AI powered smart video analytics that can simultaneously monitor for intrusions, verify identity, and analyze human activity at the workplace.

The system leverages deep learning models such as Convolutional Neural Networks (CNNs) and pose/activity recognition frameworks to:
- Detect presence and motion.
- Recognize faces for attendance tracking.
- Classify activity states (e.g., typing, walking, idle, sleeping).
- Automatically notify managers when an employee is not working or when an intruder enters restricted zones.

This leads to a more secure, automated, and productive environment, reducing the need for manual monitoring and intervention.

5. Related Work

a)Sensor-Based Systems:
Earlier systems used PIR (Passive Infrared), ultrasonic, or vibration sensors to detect motion. These were cost-effective but lacked intelligence — unable to differentiate between humans, animals, or moving objects — leading to frequent false alarms.

b) Conventional CCTV Surveillance:
Traditional CCTV systems continuously record footage and depend on human vigilance. This

makes them prone to fatigue-based errors and delayed responses, as operators may overlook incidents due to long monitoring hours.

c) Raspberry Pi–Based Smart Surveillance:
Recent projects used Raspberry Pi and OpenCV for real-time face recognition and alert generation. While effective for small setups, such systems face scalability and processing limitations for large workplaces.

d) AI-Based Intelligent CCTV:
Machine learning methods using Haar Cascades, SVM, and CNN models improved detection accuracy. These systems reduced false positives and introduced real-time tracking but were primarily focused on intruder detection, not workplace behavior.

e) Workplace Activity Monitoring (New Addition):
Recent advancements use pose estimation (e.g., MediaPipe, OpenPose) and deep neural networks to analyze human actions. Such systems can classify whether an employee is actively working, idle, or absent. Integrating this with face recognition creates a holistic workplace monitoring platform that addresses both security and productivity.

6. Methodology

Of course. Here are the key points of your methodology, summarized in a concise format suitable for a research paper, based on the chart you provided.

1. Data Collection & Pre-processing

- The dataset was developed for an **intruder detection** and **human activity recognition** system.
- The data was partitioned into an **85% training set** and a **15% validation set**.
- Data augmentations were applied to improve model generalization, including **resizing** images to 128x128, random **flipping**, **rotation**, and **normalization**.
- The core of the system utilizes **face recognition** technology.

2. Model Architecture

- A custom Convolutional Neural Network (CNN), designated **SmallCNN**, was architected for this study.
- The model features **three convolutional blocks**, each following a `Conv -> BatchNorm -> MaxPool` sequence for feature extraction.
- A **classifier head** consisting of a `Fully Connected -> Dropout -> Output` sequence is used for final prediction.

3. Training Configuration

- The network weights were optimized using the **AdamW optimizer**.
- **CrossEntropyLoss** was selected as the objective function for training.

- The model was trained for a baseline of **6 epochs**, utilizing an **early stopping** criterion with a patience of 3 to mitigate overfitting.

4. Evaluation Metrics

- Quantitative performance was assessed using **Accuracy**, **Precision**, **Recall**, and **F1-score** (both **macro** and **weighted** averages).
- Qualitative analysis of classification performance was conducted using **Confusion Matrices**.

7. Result

The proposed system was implemented and its core modules were evaluated on their respective validation datasets. The quantitative results for each model demonstrate the system's capabilities in face recognition, human activity analysis, and intruder detection. The detailed performance metrics are presented below.

1. Face Recognition Module

The face recognition model, tested on a dataset of **1340** images across **3** distinct classes, achieved a validation accuracy of **79.60%**. The performance metrics, using weighted averaging to account for class distribution, were a **precision of 80.36%**, a **recall of 79.60%**, and an **F1-score of 78.92%**. These results indicate a strong and balanced performance in correctly identifying enrolled individuals.

2. Human Activity Recognition Module

The human activity model was evaluated on the largest dataset, comprising **7574** images across **2** classes (e.g., 'working' and 'idle'). The model achieved an overall validation accuracy of **72.89%**.

The weighted F1-score was **61.46%**. It is important to note that the macro-averaged F1-score was significantly lower at **42.16%**. This discrepancy suggests a notable class imbalance within the human activity dataset, where the model performs more effectively on the majority class.

3. Intruder Detection Module

The intruder detection model, evaluated on a dataset of **998** images with **2** classes (e.g., 'authorized' and 'intruder'), achieved a validation accuracy of **77.85%**. The model's weighted metrics showed a **precision of 72.84%**, a **recall of 72.48%**, and an **F1-score of 72.66%**. This demonstrates the model's solid capability to effectively distinguish between authorized and unauthorized individuals in the monitored environment.

8. Conclusion

This research developed and evaluated **Contexa Vision**, an AI-powered system integrating **intruder detection**, **face recognition for attendance**, and **human activity monitoring**. Experimental results validated the system's effectiveness, with the face recognition and intruder detection modules achieving promising accuracies of **79.60%** and **77.85%**, respectively. While the human activity module showed potential, its performance indicated challenges with class imbalance in the dataset that require further work. Future enhancements will focus on **integrating with IoT sensors** and **exploring edge AI** to improve privacy and functionality. In conclusion, Contexa Vision provides a robust framework for an intelligent surveillance system that can significantly improve both workplace security and operational efficiency.

8. References

1. Deng, J., Guo, J., & Xue, N. (2019). *ArcFace: Additive Angular Margin Loss for Deep Face Recognition.* In **Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)** (pp. 4690–4699).
2. Schroff, F., Kalenichenko, D., & Philbin, J. (2015). *FaceNet: A Unified Embedding for Face Recognition and Clustering.* In **Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)** (pp. 815–823).
3. He, K., Zhang, X., Ren, S., & Sun, J. (2016). *Deep Residual Learning for Image Recognition.* In **Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)** (pp. 770–778).
4. Simonyan, K., & Zisserman, A. (2015). *Very Deep Convolutional Networks for Large-Scale Image Recognition (VGGNet).* **International Conference on Learning Representations (ICLR)**.
5. Redmon, J., & Farhadi, A. (2018). *YOLOv3: An Incremental Improvement.* **arXiv preprint arXiv:1804.02767.**
6. Ren, S., He, K., Girshick, R., & Sun, J. (2015). *Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks.* **Advances in Neural Information Processing Systems (NeurIPS)**, 28.
7. Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., & Berg, A. C. (2016). *SSD: Single Shot MultiBox Detector.* In **European Conference on Computer Vision (ECCV)** (pp. 21–37). Springer.
8. Tran, D., Bourdev, L., Fergus, R., Torresani, L., & Paluri, M. (2015). *Learning Spatiotemporal Features with 3D Convolutional Networks (C3D).* In **Proceedings of the IEEE International Conference on Computer Vision (ICCV)** (pp. 4489–4497).
9. Carreira, J., & Zisserman, A. (2017). *Quo Vadis, Action Recognition? A New Model and the Kinetics Dataset (I3D).* In **Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)** (pp. 6299–6308).
10. Tan, M., & Le, Q. V. (2019). *EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks.* In **Proceedings of the International Conference on Machine Learning (ICML)** (pp. 6105–6114).
11. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., ... & Houlsby, N. (2021). *An Image Is Worth 16×16 Words: Transformers for Image Recognition at Scale (ViT).* **International Conference on Learning Representations (ICLR)**.
12. Wang, H., Wang, Y., Zhou, Z., Ji, X., Gong, D., Zhou, J., ... & Liu, W. (2018). *CosFace: Large Margin Cosine Loss for Deep Face Recognition.* In **Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)** (pp. 5265–5274).