الجمهوريــة الجزائريــة الديمقراطيــة الشعبيــة

People's Democratic Republic of Algeria

و البـحـث الـعـلـمـي وزارة الـتـعـلـيـم الـعـالـي

Ministry of Higher Education and Scientific Research

Mustapha Stambouli University

Mascara

جامعة مصطفى اسطمبولي

معسكر

Faculty of Exact Sciences

Computer Science Department

كلية العلوم الدقيقة

قسم الاعلام الالي

DOCTORALTHESIS

A thesis submitted in partial fulfilment of the requirements for the 3rd cycle (LMD) PhD degree in informatics

Option : Artificial Intelligence and Its Applications

Entitled:

Visual recognition for IoT-based smart city surveillance

Presented by*: *Medjdoubi Abdelkader

On : 28/05/2024

In front of the jury:

President	Mr. Bachir Bouiadjra Rochdi	MCA	University of Mascara
Examiner	Mr. Kechar Bouabdellah	Professor	University of Oran
Examiner	Mr. Rebbah Mohamed	Professor	University of Mascara
Examiner	Mr. Zaagane Mohamed	MCA	University of Mascara
Supervisor	Mme. Meddeber Meriem	MCA	University of Mascara
Guest	Mme. Yahyaoui Khadidja	Professor	University of Mascara

Academic Year: 2023-2024

الجمهوريـــة الجزائريـــة الديمقراطيـــة الشعبيـــة

République Algérienne Démocratique et Populaire

و البـحـث الـعـلـمــي وزارة الـتـعـلــيــم الـعـالي

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Université MUSTAPHA Stambouli

Mascara

جامعة مصطفى اسطمبولي

معسكر

Faculté des Sciences Exactes

Département d'informatique

كلية العلوم الدقيقة

قسم الاعلام الالي

THESE de DOCTORAT

Spécialité : L'intelligence artificielle et ses applications

Intitulée :

Reconnaissance visuelle pour la surveillance des villes intelligents basée sur l'IoT

Présentée par* : *Medjdoubi Abdelkader

Le : 28/05/2024

Devant le jury :

Président	Mr. Bachir Bouiadjra Rochdi	MCA	Université de Mascara
Examinateur	Mr. Kechar Bouabdellah	Professeur	Université de Oran
Examinateur	Mr. Rebbah Mohamed	Professeur	Université de Mascara
Examinateur	Mr. Zaagane Mohamed	MCA	Université de Mascara
Encadreur	Mme. Meddeber Meriem	MCA	Université de Mascara
Invité	Mme. Yahyaoui Khadidja	Professeur	Université de Mascara

Année Universitaire : 2023-2024

To me and to whom stood by me

Dedication

First and foremost, I am profoundly thankful to God for making me believe, dream, and guide me through the road until the goal becomes a reality·

To my parents, God bless them for their unwavering support, encouragement, and prayers throughout this challenging yet rewarding pursuit· Your spiritual guidance has been a source of strength, and your belief in my abilities has been a constant motivation·

Special thanks are due to my dedicated teachers and mentors· To Meddeber Meriem and Yahyaoui Khadidja· Your guidance, expertise, and constructive feedback have shaped my intellectual growth and contributed significantly to the quality of my research· I am fortunate to have had the privilege of learning from such esteemed individuals· It must be said that I was so lucky to have a supervisor like yours, Madam Meriem·

To my esteemed teachers from primary school, Mme Chrifa, my teachers and mentors, mister Ben Ammara and Aid Benzarga from high school, and Dr· Kada BenYahia and Dr· Boudia Mohamed Amine from Saida University, you are the architects of my intellect, who transformed curiosity into knowledge· Your wisdom ignited a passion for learning, and your guidance sculpted the path to this academic summit· Thank you for not just imparting information but also for being catalysts of inspiration and mentors of distinction· Your impact is woven into the fabric of my achievement, and for that, I am forever grateful·

I extend my appreciation to my friends and family who stood by me during the highs and lows of this academic endeavor·

To Abddeli Wafaa, who was supportive from the beginning of the journey, I am really grateful for life· Thank you for your patience and being there for me·

Acknowledgements

I extend my heartfelt gratitude to the members of the jury for their invaluable contributions during the discussions on both my dissertation and doctoral thesis. Mr. Bachir Bouiadjra Rochdi, Mr. Kechar Bouabdellah, Mr. Zaagane Mohamed and Mr. Rebbah Mohamed, your thoughtful insights and constructive feedback have and will play a crucial role in refining and enhancing the depth of my research. I deeply appreciate the time and expertise each of you dedicated to the evaluation process, enriching the scholarly dialogue.

I am grateful for the privilege of engaging with such distinguished scholars, and I look forward to incorporating your valuable suggestions into my ongoing and future research endeavors.

.

Abstract

The rapid evolution of edge technology and Internet of things technology, combined with sophisticated facial recognition systems, has catalyzed a transformative paradigm in urban surveillance, culminating in the emergence of smart cities. This thesis delves into the intersectionality of Internet of things and facial recognition methodologies within the urban milieu, examining their synergistic integration in the construction of intelligent, responsive, and secure urban environments. The exploration encompasses a comprehensive analysis of the technological intricacies inherent in Internet of things sensor networks and facial recognition algorithms, while critically evaluating their ethical, legal, and societal implications. The dissertation provides an exhaustive overview of the deployment of Internet of things devices, spanning from embedded sensors in urban infrastructure to wearable gadgets, facilitating real-time data acquisition and analysis. When coupled with facial recognition technologies, this data forms the fundamental backbone of smart city surveillance, presenting distinctive opportunities for urban management, security enhancement, and resource optimization. This research meticulously scrutinizes challenges related to data privacy and algorithmic biases, proffering innovative solutions and policy frameworks to effectively address these concerns. Moreover, the study investigates the integration of edge computing, optimizing data processing and analysis for resource-constrained IoT devices. In addition to critical examination, attention is directed towards recent advancements in machine learning and artificial intelligence, elevating the accuracy, resilience, and real-time processing capabilities of facial recognition systems internet of things and edge technology based solutions. Showing a theoretical understanding and practical implementation of smart city surveillance real world application.

Keywords: Artificial Intelligence, Computer Vision, Deep Learning, Edge Technology, Face Recognition, Internet of Things, Smart City

Résumé

L'évolution rapide de la technologie de pointe et de la technologie de l'Internet des objets, combinée à des systèmes sophistiqués de reconnaissance faciale, a catalysé un paradigme de transformation de la surveillance urbaine, aboutissant à l'émergence de villes intelligentes. Cette thèse explore l'intersectionnalité des méthodologies de l'Internet des objets et de la reconnaissance faciale au sein du milieu urbain, en examinant leur intégration synergique dans la construction d'environnements urbains intelligents, réactifs et sécurisés. L'exploration englobe une analyse complète des subtilités technologiques inhérentes aux réseaux de capteurs de l'Internet des objets et aux algorithmes de reconnaissance faciale, tout en évaluant de manière critique leurs implications éthiques, juridiques et sociétales. La thèse fournit un aperçu exhaustif du déploiement des dispositifs de l'Internet des objets, allant des capteurs intégrés dans les infrastructures urbaines aux gadgets portables, facilitant l'acquisition et l'analyse de données en temps réel. Associées aux technologies de reconnaissance faciale, ces données constituent l'épine dorsale fondamentale de la surveillance des villes intelligentes, offrant des opportunités distinctes pour la gestion urbaine, l'amélioration de la sécurité et l'optimisation des ressources. Cette recherche examine méticuleusement les défis liés à la confidentialité des données et aux biais algorithmiques, proposant des solutions innovantes et des cadres politiques pour répondre efficacement à ces préoccupations. De plus, l'étude étudie l'intégration de l'informatique de périphérie, l'optimisation du traitement et de l'analyse des données pour les appareils IoT à ressources limitées. En plus de l'examen critique, l'attention est dirigée vers les progrès récents de l'apprentissage automatique et de l'intelligence artificielle, améliorant la précision, la résilience et les capacités de traitement en temps réel des systèmes de reconnaissance faciale internet des objets et des solutions basées sur la technologie de pointe. Montrant une compréhension théorique et une mise en œuvre pratique de l'application réelle de la surveillance des villes intelligentes.

Mots-clés: Intelligence Artificielle, Computer Vision, Deep Learning, Technologie de pointe, Reconnaissance Faciale, Internet des objets, Ville intelligente.

ملخص

أدى التطور السريع للتكنولوجيا المتطورة وتكنولوجيا إنترنت الأشياء ، جنبا إلى جنب مع أنظمة التعرف على الوجه المتطورة ، نموذجا تحويليا في المراقبة الحضرية ، وبلغت ذروتها في ظهور المدن الذكية. تتعمق هذه الأطروحة في تقاطع إنترنت الأشياء ومنهجيات التعرف على الوجه داخل البيئة الحضرية ، وتدرس تكاملها التآزري في بناء بيئات حضرية ذكية وسريعة الاستجابة وآمنة. يشمل الاستكشاف تحليلا شاملا للتعقيدات التكنولوجية المتأصلة في شبكات استشعار إنترنت الأشياء وخوارزميات التعرف على الوجه ، مع تقييم نقدي لآثارها الأخلاقية والقانونية والمجتمعية. تقدم الرسالة نظرة عامة شاملة على نشر أجهزة إنترنت الأشياء ، والتي تمتد من أجهزة الاستشعار المدمجة في البنية التحتية الحضرية إلى الأدوات القابلة للارتداء ، مما يسهل الحصول على البيانات وتحليلها في الوقت الفعلي. عندما تقترن هذه البيانات بتقنيات التعرف على الوجه ، فإنها تشكل العمود الفقري الأساسي لمراقبة المدينة الذكية ، مما يوفر فرصا مميزة للإدارة الحضرية ، وتعزيز الأمن ، وتحسين الموارد. يفحص هذا البحث بدقة التحديات المتعلقة بخصوصية البيانات والتحيزات الخوارزمية ، ويقدم حلولا مبتكرة وأطرا سياسية لمعالجة هذه المخاوف بشكل فعال. علاوة على ذلك ، تبحث الدراسة في تكامل الحوسبة المتطورة ، وتحسين معالجة البيانات وتحليلها لأجهزة إنترنت الأشياء المحدودة الموارد. بالإضافة إلى الفحص النقدي ، يتم توجيه الانتباه نحو التطورات الحديثة في التعلم الآلي والذكاء الاصطناعي ، ورفع الدقة والمرونة وقدرات المعالجة في الوقت الفعلي لأنظمة التعرف على الوجه إنترنت الأشياء والحلول القائمة على التكنولوجيا المتطورة. عرض الفهم النظري والتنفيذ العملي للمراقبة المدينة الذكية تطبيق العالم الحقيقي.

الكلمات المفتاحية: الذكاء الاصطناعي ، رؤية الكمبيوتر ، التعلم العميق ، تكنولوجيا الحافة ،
التعرف على الوجوه ، إنترنت الأشياء ، المدينة الذكية

Content

List of Figures

List of Tables

List of Abbreviations

General Introduction

In the rapidly evolving landscape of technology, the fusion of the Internet of Things (IoT), edge technology (ET), and advanced artificial intelligence (AI) techniques such as computer vision (CV) and deep learning (DL) has ushered in a new era of innovation and efficiency. Smart cities (SC), characterized by interconnected devices, sensors, and intelligent systems, represent the pinnacle of this technological revolution. These cities leverage IoT and ET to enhance the quality of urban life, optimize resource management, and improve overall safety and security. At the heart of these SCs lies an intricate web of surveillance systems powered by cutting-edge technologies. Surveillance in SCs extends far beyond conventional methods; it now integrates sophisticated CV algorithms, enabling real-time analysis of vast streams of data. This amalgamation of IoT, ET, and CV not only provides unprecedented insights into urban dynamics but also forms the backbone of numerous applications, ranging from traffic management and environmental monitoring to public safety and healthcare [1].

However, the realization of these SC monitoring systems faces major problems. The sheer volume of data generated by IoT devices and surveillance cameras strains traditional centralized computing systems, leading to latency, bandwidth issues, and privacy concerns. It is at this critical juncture that ET emerges as a game-changer. ET allows for data processing closer to the source, reducing latency, enhancing responsiveness, and ensuring data privacy [1][2].

Furthermore, the synergy between ET and DL algorithms has offered new opportunities for intelligent data analysis. DL, a subtype of AI, empowers machines to learn from huge datasets, enabling them to spot patterns, make judgments, and even predict future occurrences. In the context of SC surveillance, DL algorithms play a vital role in decoding complicated visual data, finding anomalies, and boosting the entire security architecture, especially in the context of surveillance, where biometric applications are also witnessing a huge advancement. In the context of biometric technology, an important field in the sphere of digital security and identity verification, has altered the way we interact with the digital world. Among the different biometric modalities, face recognition (FR) stands out as a significant and widely utilized approach, flawlessly mixing ease with powerful security measures [2][3]. This disruptive technology finds applications across numerous areas, altering industries and boosting user experiences. FR technology acts as a strong tool for enhancing security measures. Traditional authentication mechanisms like passwords and PINs are more vulnerable to breaches, leading to an increase in identity theft and unauthorized access events. FR, with its capacity to assess unique facial traits, presents a highly secure alternative. From safeguarding cellphones and laptops to restricting access to sensitive places within enterprises, this technology guarantees that only authorized personnel gain entry, increasing overall security standards. Biometric applications, particularly FR, have crossed traditional limits, invading various areas of our lives [2][3]. By seamlessly merging security, efficiency, and personalized experiences, this technology continues to transform industries and pave the way for a future where digital interactions are secure, convenient, and suited to individual needs. As research and development in the field of biometrics expand, the possible applications of these technologies are infinite, mainly integrating ET, IoT, and DL approaches, offering a future where security and convenience happily coexist. Throughout this study, our efforts were focused on putting answers into practice and resolving the problems that come with surveillance technologies. Strong features from multiple state-of-the-art technologies were combined to create a clever real-world application. This application is based on theoretical study, which emphasizes how important it is to provide solutions in the current environment. This study's entire strategy is implemented in stages, starting with careful data collection, moving through complex processing and manipulation, and ending with advanced categorization methods. Achieving high-performance results with optimal accuracy was the ultimate goal, all while adhering to real-time responsiveness time. This multidimensional approach emphasizes the necessity of such solutions in the current socio-technological environment in addition to reflecting the theoretical foundations' practical implementation.

1.1 Problematic

Creating a privacy-aware and ethical surveillance framework utilizing advanced technologies such as ET, IoT and FR has become essential for improving security. The widespread use of surveillance technologies in SC, however, presents serious ethical and privacy issues from one side and the effectiveness of real-time response and performance outcomes from the other, which calls for careful investigation and creative solutions. In an effort to solve these issues, this study puts forth a thorough framework that prioritizes ethical issues and protects people's right to privacy while also optimizing the effectiveness of SC monitoring. The principal elements of this framework comprise:

- **IoT and Edge Integration:** Investigating how to combine IoT devices and ET to maximize the benefits of centralized data processing while also improving the effectiveness of surveillance systems.
- **AI and CV Algorithms:** An assessment of how these technologies affect the impartiality, accuracy, and potential misuse of surveillance techniques, along with suggestions for how to address these problems and encourage responsible use.
- **Privacy Preservation:** Investigating methods and procedures that guarantee citizens' privacy to be maintained throughout large-scale data gathering and processing, with a particular emphasis on the moral application of FR technology.

By offering a comprehensive and morally sound perspective on surveillance, this research hopes to add to the growing conversation about SC development. It recognizes the critical role that technology plays in urban management while also upholding residents' rights and concerns.

1.2 Thesis Scope and Objectives

This comprehensive thesis thoroughly investigates the intricate intersection of IoT, ET, and CV within the specific context of SC surveillance. The central focus lies in unraveling the multifaceted challenges and promising opportunities that emerge as a result of integrating these cutting-edge technologies. Through an in-depth analysis of real-world implementations and compelling case studies, this research aspires to shed light on the efficacy of ET in managing the overwhelming influx of data generated by IoT devices and surveillance cameras. Moreover, this thesis aims to meticulously probe into the influence of contemporary CV techniques and advanced DL algorithms on the precision and efficiency of SC surveillance systems. Additionally, exploring their pervasive effects on our daily lives across diverse domains by delving into the intricacies of these technologies, this study seeks to understand their transformative impact on the landscape of urban surveillance, thereby contributing valuable insights to the fields of security and public safety. The advent of AI has ushered in a wave of innovation, enabling the development of technologies that profoundly impact various aspects of society. This thesis endeavors to unravel the nuances of these advancements, examining their implications and the ways in which they shape our interactions with the world. Through rigorous analysis and empirical exploration and by going into these complexities, this research contributes significantly to the ongoing discourse surrounding the evolution of surveillance systems and the transformative potential of emerging technologies in shaping the future of our cities and societies.

The following list, which is presented in chronological order, demonstrates the various scientific contributions that came as a result of the study for this thesis, including two communications and a journal publication:

- **Communication:** Abdelkader Medjdoubi, Meriem Meddeber and Khadidja Yahyaoui, " Deep Learning IoT Based Smart Face Recognition System", presented virtually at the International Conference on Artificial Intelligence and Soft Computing, (ICAISC), Krakow,Poland, Dec, 2022.
- **Communication:** Abdelkader Medjdoubi, Meriem Meddeber and Khadidja Yahyaoui, "Face Recognition Surveillance System IoT and Deep Learning Based" presented virtually at the International Conference on Engineering & Technology, (ICET), Salem,India, Oct, 2023.
- **Publication:** Abdelkader Medjdoubi, Meriem Meddeber and Khadidja Yahyaoui, "Smart City Surveillance: Edge Technology Face Recognition Robot Deep Learning Based", International Journal of Engineering, Oct, 2023.

1.3 Thesis Structure

This thesis is structured around four chapters, including the above introduction and a conclusion. In the following, we provide a short description of the subjects treated in the subsequent chapters:

- **Chapter 1 - Computer Vision in Surveillance Systems and Smart City:** This chapter goes through the realm of CV and its pivotal significance within surveillance systems. It comprehensively explores the fundamental tasks essential for security systems, encompassing object detection, object tracking, behavior analysis, and anomaly detection. The discourse culminates with an insightful examination of the intersection between the aspect of security and the domain of CV, offering valuable insights into the intricate relationship between these critical components in the realm of security studies. Also an exhaustive account of the SC concept is provided, encompassing its definition, constituent elements, distinctive features, and inherent challenges. Subsequently, the discussion transitions to the domain of SC security, dig into the strategic integration of surveillance techniques to establish and fortify these essential facets within SCs.
- **Chapter 2 - Intelligent Surveillance Systems:** This chapter focuses on the development of intelligent security solutions in the world of surveillance, focusing on the contemporary used technologies. These technologies encompass ET, Edge Intelligence, Edge Artificial Intelligence, and Artificial Intelligence of Things. It aims to elucidate the construction of smart security solutions by providing comprehensive insights of its components, including integration of IoT micro-controllers. Furthermore, this part explores real-world applications of these solutions, shedding light on their pervasive impact on various facets of our daily lives.
- **Chapter 3 - Facial Recognition for Smart City Surveillance:** This chapter critically analyzes FR systems, covering aspects ranging from their fundamental definition to the complex techniques and benchmarking datasets. Additionally, a comprehensive review of existing works in the literature is presented.
- **Chapter 4 – Our Contribution:** The presented works and methodologies in the previous chapters serves as a foundation for the subsequent discussion of our unique contributions, wherein we detail the techniques, approaches, and the obtained results in our research endeavors.
- **Conclusion and Future Work:** This section provides a concise summary of the primary objectives and key topics explored throughout the thesis. Additionally, it addresses the prospective areas for future research and outlines potential perspectives in the field.

Chapter 01 : Computer Vision in Surveillance Systems and Smart Cities

1.1 Introduction

In an era of extraordinary technological developments, CV stands as a beacon of innovation, altering the way we see, comprehend, and interact with the visual world. CV is a multidisciplinary field that seeks to enable robots to interpret and make sense of visual information in the same way as the human visual system does. Consider a future in which machines not only "see" images and movies but also comprehend their substance, context, and significance. This field of CV takes through a panorama of algorithms and image processing techniques, all of which contribute to the emergence of computers with visual perception. The CV domain has a huge impact that influences the future of technology and ushers in an era where machines genuinely perceive the visual richness of our environment as we travel this exciting landscape [2][3].

1.1.1 Computer Vision a Defintion

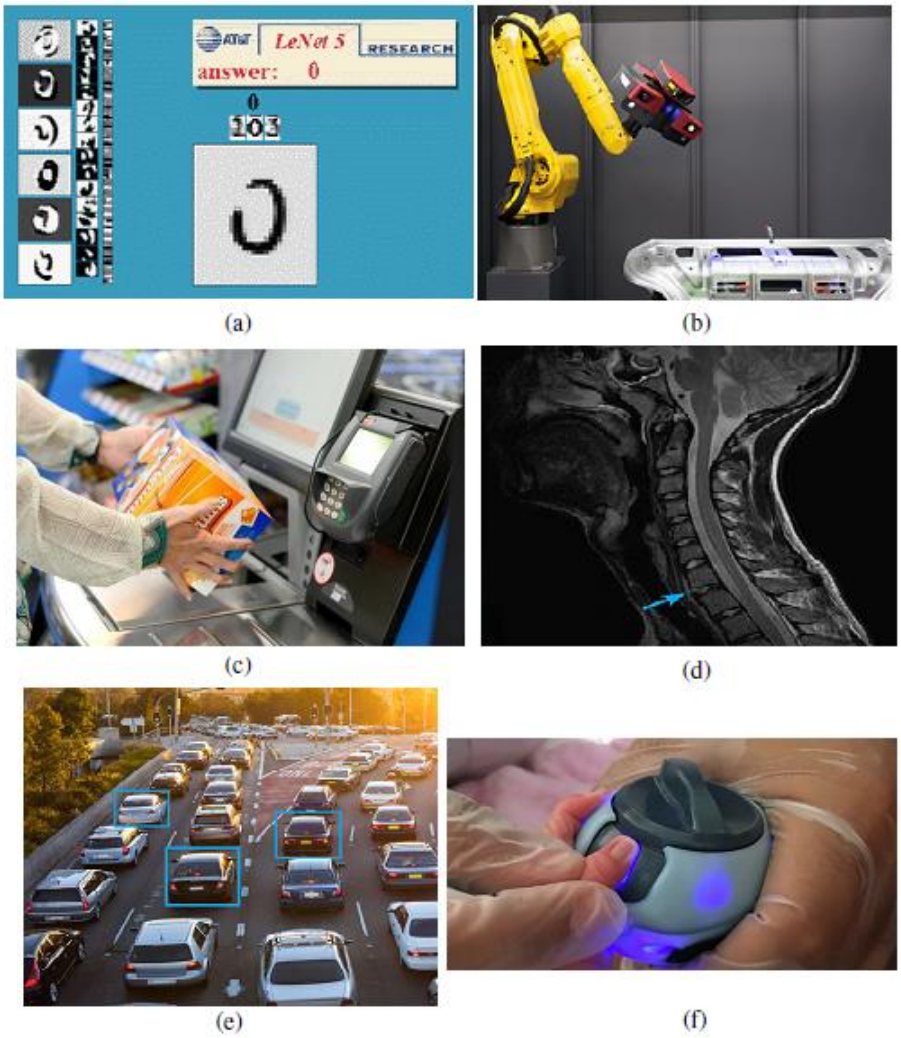
At the intersection of computer science, AI, and image processing, CV is an exciting and quickly developing topic. It includes the research, creation, and use of algorithms and methods that allow machines to take in, comprehend, and interpret visual data from the environment. CV, which draws its inspiration from the human visual system, aims to equal or perhaps surpass our capacity for object recognition, scene comprehension, and the inference of significant information from photos and videos. It is a multidisciplinary scientific area that uses computers to thoroughly analyze visual data, utilizing a methodology that is comparable to that of human visual systems. This area falls under the umbrella of AI and its applications [1][4].

Thanks to the widespread use of sophisticated hardware, an abundance of visual data, and developments in DL and machine learning (ML) approaches, the area of CV has made considerable strides in recent years. These advancements have made it possible for CV applications to go beyond traditional fields like robotics and surveillance to include various domains, including FR, medical imaging, autonomous vehicles, augmented reality, and more. CV scientists and engineers are constantly coming up with new solutions to solve difficult problems, including scene understanding, image segmentation, position estimation, and action recognition. Their efforts play a critical role in determining the course of numerous businesses, enhancing human-computer interactions, and meeting societal requirements in industries including healthcare, agriculture, transportation, and entertainment [3].

Researchers and practitioners in this vibrant and fast-paced discipline of CV are on a mission to fully exploit visual data and give machines the capacity to see the world with human-like understanding. The influence of this domain on our daily lives will inevitably increase as technology develops, heralding a time when visual intelligence will play a crucial role in our digital ecology. Today, CV is employed in a huge range of real-world applications [4][5], such as:

- **Optical character recognition :** automatically recognizing license plate numbers and reading handwritten postal codes on letters (Figure 1.1.a);

- **Machine inspection:** fast parts inspection for quality control utilizing stereo vision and specialized lighting to assess tolerances on auto body or aircraft wing parts (Figure 1.1.b) or using X-ray vision to search for flaws in steel castings;
- **Retail:** recognizing objects for automated checkout lanes (Figure 1.1.c);
- **3D model building photogrammetry :** totally automated creation of 3D models from aerial images for usage in applications of 3D printing;
- **Medical imaging:** logging pre- and post-operative imaging (Figure 1.1.d) or carrying out long-term investigations of aging brain morphology in patients;
- **Match move:** by tracking feature points in the source video, it is possible to estimate the 3D camera movements and the form of the surroundings when combining computer-generated imagery with real-world footage. These methods are frequently employed in the Hollywood filmmaking sector. (e.g., in movies such as Jurassic Park (1993) and in The Irishman (2019));
- **Motion capture (mocap):** capturing actors for computer animation by employing retro-reflective markers viewed from numerous cameras or other vision-based techniques;
- **Surveillance:** monitoring for intruders, examining traffic on the highway (Figure 1.1.e), video surveillance in public spaces;
- **Fingerprint recognition and biometrics:** for forensic applications as well as automatic access authentication (Figure 1.1.f).



(a) (b) (c) (d) (e) (f)

Figure 1.1: Some industrial applications of computer vision: (a) optical character recognition (OCR) http://yann.lecun.com/exdb/lenet/; (b) mechanical inspection https://krtkl.com/industries/vision/; (c) retail https://www.ocregister.com/; (d) medical imaging https://www.melbourneradiology.com.au; (e) Surveillance https://www.iteslab.com; (f) Fingerprint recognition and biometrics https://www.nature.com/ [6][7][8];

1.2 Computer Vision for Surveillance

One of the most influential areas where CV has had a profound impact is in the field of surveillance. As a crucial component of security and public safety, surveillance has historically relied on human operators to watch and interpret video feeds from cameras installed in various surroundings. However, this human-centered strategy has built-in drawbacks like subjectivity, weariness, and the difficulty of comprehending massive volumes of data in real-time [4]. CV can be a game-changer in this situation. It gives surveillance systems the ability to automatically recognize, track, and analyze objects, actions, and patterns inside video streams by utilizing the power of advanced algorithms and ML approaches. These intelligent solutions considerably improve the precision, effectiveness, and scalability of surveillance operations while also lightening the workload for human operators [4][5]. The use of CV in surveillance are numerous and diverse. Its techniques provide a complete ability for handling the complex difficulties of contemporary security and surveillance, from identifying people and vehicles to observing crowd behavior and spotting anomalies. This domain allows for the potential to turn passive cameras into active, intelligent agents that contribute dynamically to the structure of a SC, whether it be for spotting potential security threats, guaranteeing public safety during significant events, or optimizing traffic management in busy urban centers [9]. It is abundantly evident that CV has made significant contributions to this domain by integrating new technologies to usher in a new era of efficiency and accuracy to create safer places [10].

We shall dive into the complexity of numerous CV techniques designed specifically for surveillance throughout this chapter. We'll look at how object detection (OD) and object tracking (OT) algorithms make it possible to automatically identify and keep an eye on interesting objects. Additionally, we'll examine behavior analysis (BA) and anomaly detection (AD) to highlight how intelligent systems can identify strange behaviors and patterns in visual data. Also, we will go through FR and identification, enabling surveillance technologies to identify people in crowds, which will be presented in the fourth chapter.

1.3 Object Detection and Tracking

In a variety of industries, including not only CV but also robotics, autonomous driving, surveillance, healthcare, sports analytics, and more, object identification and tracking have taken on increasing importance. While tracking entails observing the movement of items that have been recognized over time, OD aims to locate objects within an image or video sequence. In this part, we will examine the fundamentals of both tasks as well as the most recent developments and applications.

1.3.1 Defintion of Object Detection and Tracking

OD is a fundamental task in CV; it involves identifying objects within an image or video sequence and locating the bounding box (or region) around each object, along with classifying the object into a specific category (e.g., person, car, dog). Accurately identifying objects among differences in scale, orientation, lighting, and occlusions are the main challenges [11]. OD can be performed using various methods, previously with traditional techniques that relied heavily on handcrafted features and complex algorithms, but nowadays including DL methods like YOLO (You Only Look Once), SSD (Single Shot Detector), and Faster R-CNN (Region-based Convolutional Neural Networks) [12]. These approaches use convolutional neural networks (CNN) to learn features from images and detect objects based on their appearance, texture, and other attributes. An example of an OD application is provided in Figure 1.2. Based on the YOLO DL model integrated inside a car camera on the streets.

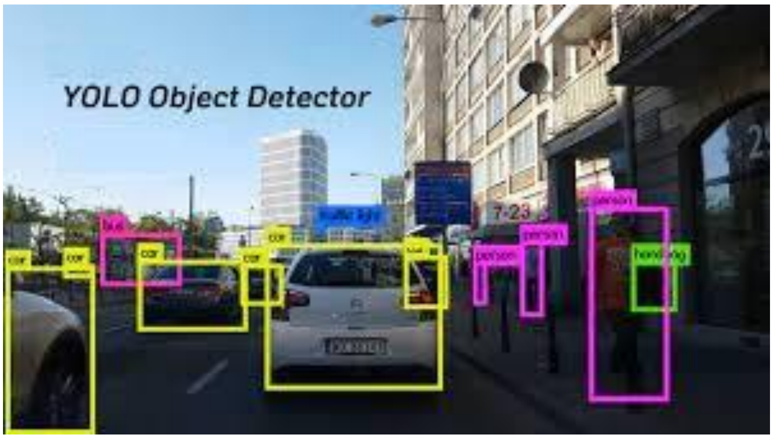


Fig. 1.2: An illustration of an OD application based on the YOLO model [13].

While OD focuses on identifying objects within individual frames, OT extends this concept by following the detected objects over time across consecutive frames in a video sequence. By creating links between objects in various frames, OT algorithms make it possible to track an object's movement across time even when it moves, rotates, or changes appearance. [11][12]. This is especially useful in situations like surveillance, sports analysis, and robotics where the target objects are moving. Various techniques and methods have been used to achieve this task, including advanced DL and ML methods taking into consideration complicated settings and enhancing tracking accuracy like DeepSORT and GOTURN (Generic Object Tracking Using Regression Networks), as well as traditional approaches like Mean Shift, Kalman Filters and particle filters [11][12] [14]. An example of an OT application is provided in Figure 1.3. Based on the DeepSORT DL model integrated inside a live streaming camera that follows players movements around the field.

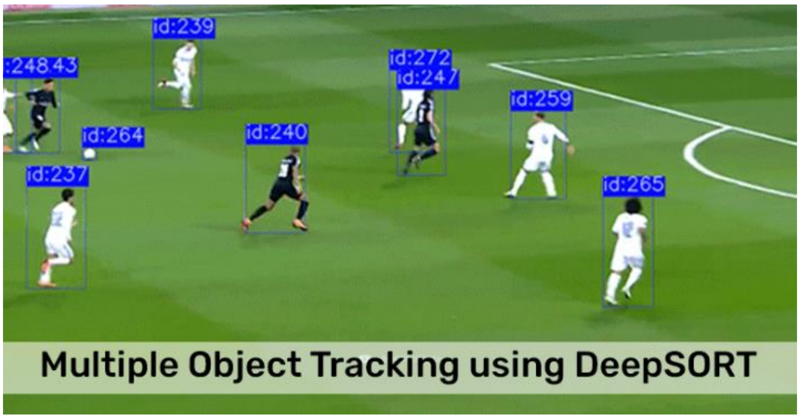


Fig. 1.3: An illustration of an OT application based on the DeepSORT model [14].

1.3.2 Advancements in Object Detection and Tracking

OD and OT are highly active fields. Both have seen significant advancements due to the rapid progress in DL, CV, and related technologies. Below, we list some key advancements that were prominent up to that point:

- **Real-time Performance:** The emphasis on instantaneous OT and OD has been one of the major developments in recent years. Modern algorithms are well suited for applications that demand quick decision-making, such as autonomous vehicles and robots, because they can operate at astonishing rates while maintaining high accuracy [8].
- **Multi-Object Tracking:** Due to potential occlusions and interactions between items, tracking many objects at once is a difficult operation. Multiple moving objects may now be tracked by sophisticated algorithms and DL models, enabling the observation and analysis of complicated scenarios. [8][9].
- **Robustness to Occlusions and Variability:** Modern models have shown improved resistance to occlusions and changes in scale, position, and lighting conditions. As a result, they are dependable even in difficult situations when objects could be partially covered or move erratically [9].

1.3.3 Intersection between Object Detection and Tracking

For more complete and accurate CV systems that can analyze and interpret the behavior of objects in video sequences, the intersection of OD and OT is crucial. The ability to follow and comprehend things as they move and interact in complicated contexts is improved by the combination of the two approaches. Merging the work together provides a more comprehensive understanding of object behavior in videos [10]. To understand even more this intersection, Figure 1.4 illustrates the main points on how the two approaches complete each other's work, as we can point out this representation in the form of three main steps as follows:

- **Initialization:** OD can be used as the starting point of OT. The tracking algorithms can begin tracking the objects that have already been detected in the first frame.
- **Re-identification:** OD can assist in re-identifying items that were lost during tracking in circumstances where objects could be momentarily obscured or altered in appearance.
- **Updating Object States:** The positions, sizes, velocities, and other states of monitored objects must be updated continually. When tracking may have wandered, OD can occasionally be utilized to re-detect objects and update their statuses.

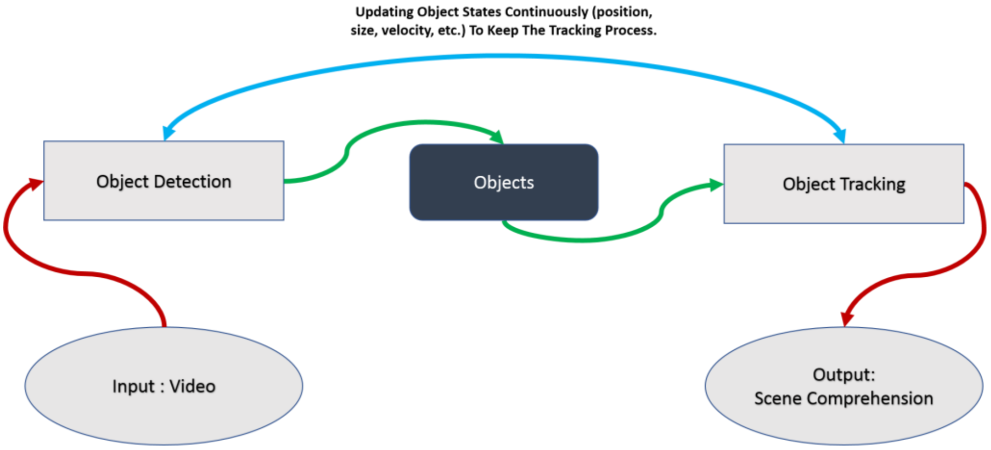

Updating Object States Continuously (position, size, velocity, etc.) To Keep The Tracking Process.
Object Detection
Objects
Object Tracking
Input : Video
Output:
Scene Comprehension

Fig. 1.4: The intersection between OD and OT .

In real-world scenarios, handling complex scenarios is a very hard task to do since objects might appear, disappear, merge, or split [10]. So, OD and OT work together to handle these complex situations, where the scene is dynamic and objects interact. Improving the overall robustness of the system by providing fresh information about the presence of new objects and tracking maintains continuity for objects that are already being monitored [10].

1.3.4 Applications of Object Detection and Tracking

Due to their capacity to automatically recognize and track objects in real-time, OD and OT technologies have a wide range of applications in numerous industries. Some of the main applications are listed below:

- **Autonomous Vehicles:** Autonomous cars must have OD and OT capabilities in order to comprehend and react to their surroundings. These tools help self-driving cars navigate more safely by helping them recognize pedestrians, vehicles, and obstructions [9].
- **Surveillance and Security:** In surveillance systems, OD and OT are crucial components that help identify intruders, suspicious activity, and unlawful access. Security systems can successfully monitor crucial regions thanks to real-time tracking [11].
- **Industrial Automation:** Applications for object identification and tracking can be found in industrial settings, including robotic automation, inventory management, and quality control on manufacturing lines. They make it possible for machines to interact with their surroundings and carry out jobs precisely [11].
- **Healthcare:** These technologies can be utilized in the healthcare industry to track medical equipment, monitor patient movement, and examine surgical techniques. They aid in enhancing patient care and streamlining hospital operations.

- **Augmented Reality :** Applications of augmented reality rely on OD and OT to smoothly overlay virtual things onto the actual world. This improves user experiences on interactive learning, entertainment, and gaming platforms [11].
- **Sports analytics:** Sports footage can be used to analyze player performance, monitor the ball, and offer real-time statistics using OD and OT systems.
- **Robotics**: Robots can interact with their environments, control items, and carry out jobs like assembly and warehouse management thanks to object identification and tracking.
- **Wildlife monitoring:** Using OD and OT, wildlife conservation operations can monitor populations, investigate animal behavior, and safeguard endangered species.

1.3.5 Challenges and Limitations of Object Detection and Tracking

Although OD and OT technologies have both advanced significantly, they still have a number of problems and restrictions. While some of these difficulties are attributable to technological limitations, others are inherent to the nature of the task. These are the main difficulties and limitations faced by both taks [9] [11]:

- **Occlusion:** Accurate detection and tracking of items can be challenging when they are partially obscured by other objects or scene elements.
- **Illumination:** The detection might be impacted by varying lighting conditions, especially in low-light settings.
- **Clutter:** Object identification and tracking systems can become confused by cluttered scenes with lots of items, which can result in false positives and losses.
- **Motion blur:** Fast-moving objects can create motion blur, which makes it difficult to detect and track them with a high accuracy.
- **Object similarity:** It can be challenging to recognize and monitor separately similar-appearing things, especially for objects that look like.
- **Real-time processing:** To support applications like autonomous driving and surveillance, object recognition and tracking must be done in real-time.
- **Greater precision:** To increase OD and OT accuracy, robustness, and speed, researchers are constantly creating new methods and structures.
- **Multi-modal sensing**: Combining data from various sensors (such as cameras, lidars, and radars) might improve the ability to track and detect objects.
- **Online learning:** Training object identification and tracking models online as opposed to offline can help them perform better and adapt to changing settings.
- **Ability to explain:** Creating clear and comprehensible object recognition and tracking models can boost confidence and comprehension in these technologies.
- **Edge computing:** By processing OD and OT close to the data source, latency can be decreased and real-time performance can be enhanced.

1.4 Behavior Analysis and Anomaly Detection

With various applications in many areas, including surveillance, healthcare, education, and entertainment, BA has grown in importance as a study field in CV. Automatic recognition and interpretation of human actions, emotions, and intents from visual data is the aim of human BA, which examines body gestures, facial expressions, postures, and bodily motions. AD, on the other hand, is concerned with locating odd or unexpected actions that depart from established patterns [11]. CV systems can deliver in-the-moment insights into human activity and quickly notify operators of potential threats or out-of-the-ordinary happenings by merging these two domains.

1.4.1 Defintion of Behavior Analysis and Anomaly Detection

The computational study and interpretation of human behaviors, interactions, and patterns utilizing a variety of methods, algorithms, and technologies is known as human BA. It involves using CV, ML, data mining, and other techniques to examine and comprehend how people behave in both virtual and physical settings. Aims to get relevant insights from many data sources, including text, photos, videos, sensor data, and so on. It involves activities like pose estimation, action recognition, gesture analysis, emotion recognition, and AD, all of which help us understand how people behave in various situations [12] [13]. The development of computing models and systems that can automatically detect, analyze, and anticipate human actions and behaviors is the ultimate goal of a BA in computer science. The ability to understand and react to human behavior is essential for developing effective and efficient technology solutions, and these models and systems will be integrated as solutions in domains like surveillance, security, healthcare, user interface design, robotics, and other related applications [14]. An example is shown in Figure 1.5, where it represents human pose estimation based on 3D poses of humans in real-time with a DL model.

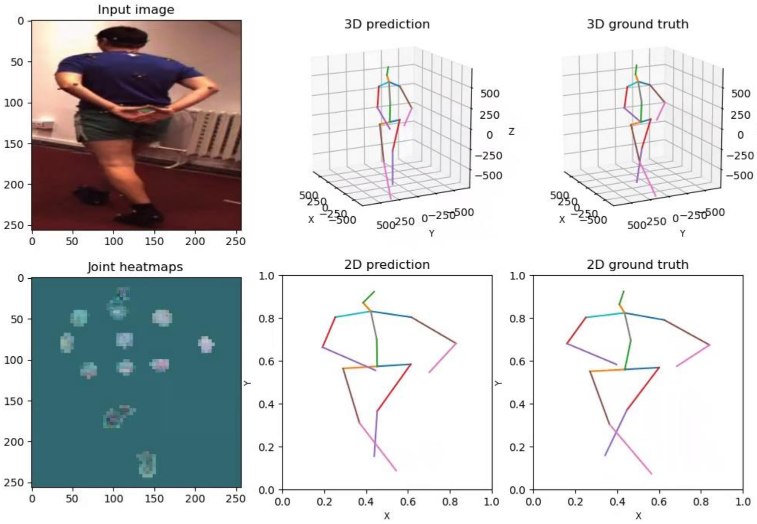


Fig. 1.5: An illustration of 3D human pose estimation DL model for human BA [15]

Finding patterns, instances, or data points that significantly depart from expected or typical behavior within a given dataset or system is referred to as AD [14]. To find anomalies or outliers that defy established patterns or trends, it includes applying a variety of algorithms, statistical techniques, and ML tools. The main goal of AD is to find strange or unexpected occurrences that may point to mistakes, fraud, flaws, potential dangers, threats, or other sorts of unexpected behavior within a system. Analyzing and observing data of many kinds, including time series data, image data, text data, and network traffic data, can be subject to AD [14] [15]. An example is represented in Figure 1.6, where it shows BA to detect anomalies. The system is based on the DL model to track people's movements inside shopping stores, extracting features and analyzing the scene to detect any suspicious customer movements.

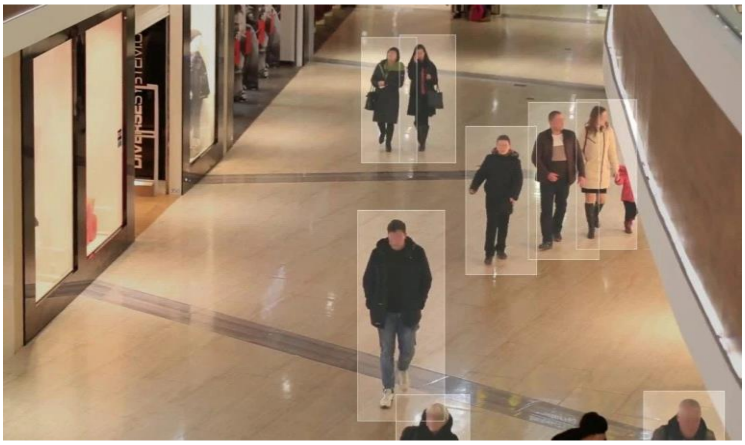


Fig. 1.6: An illustration of human behavior analysis, DL based model [9]

1.4.2 Applications of Behavior Analysis and Anomaly Detection

BA and AD have gained prominence in many fields, transforming industries and driving innovation across diverse sectors, thanks to the development of ML technologies and data analytics, revolutionizing decision-making processes, and opening the door to a deeper comprehension of both individual and group activities in a wide range of fields, including the following:

- **Healthcare and Medical Diagnostics:** Understanding human behavior is essential for precise diagnosis and efficient treatment in the field of healthcare. Healthcare workers can learn more about probable health issues or anomalies by examining patients' behaviors, such as activity patterns, sleep cycles, and communication preferences. When it comes to spotting changes in a patient's vital signs or behavior, AD algorithms can be quite helpful in enabling prompt interventions and individualized care. A typical sleep patterns, for instance, may point to underlying mental health issues, while abrupt changes in mobility may portend a forthcoming physical condition [14].
- **Cybersecurity and Fraud Detection:** The opportunities for fraud and cyber risks grow as the digital world develops. Combining BA with AD strategies provides a

powerful deterrent against such danger. Systems can quickly identify outliers that can signify unauthorized access or questionable activity by generating baseline behavior profiles for people or devices. An unexpected login location or irregular transaction behavior, for example, may set up alerts that demand immediate reactions to stop potential breaches [15].

- **Retail and Customer Experience:** Retailers are using BA more and more to improve customer experiences and increase sales. Businesses can adjust their services to meet customer preferences by closely examining purchasing trends, browsing habits, and engagement indicators. AD aids in spotting anomalies in consumer behavior, giving merchants the ability to spot possibly fraudulent activity or strange shopping trends. Additionally, merchants can improve store layouts, staffing levels, and marketing plans with the aid of real-time research of customer behavior in stores [14] and [15].
- **Transportation and Urban Planning:** Understanding human behavior is essential for maximizing traffic flow, public transportation, and infrastructure development in the fast-paced world of transportation and urban planning. Designing effective and sustainable urban systems can benefit from an analysis of commuting patterns, traffic jams, and public transportation use. Relying AD, commuter volume fluctuations can be quickly identified, allowing for prompt adjustments and interventions to reduce traffic or enhance public transit services. [14][16].
- **Workplace Productivity and Safety:** BA has a critical role to play in enhancing workplace safety and productivity. Organizations may pinpoint areas for improvement and expedite workflows by keeping track of employee habits, including task completion rates, communication preferences, and collaboration patterns. AD makes the workplace safer by identifying risky actions or deviations from established procedures. For instance, spotting odd movement patterns or equipment usage might help reduce accidents and improve workplace safety in general [16].

1.4.3 Behavior Analysis and Anomaly Detection Limitations

BA and AD are critical components in various fields and industries. However, these technologies also face several challenges and limitations. Here are some of the key points, among others:

- **Data Privacy and Ethical Concerns:** The area of data privacy and ethical considerations presents one of the main difficulties in BA and AD. Concerns about user rights and potential abuse arise when collecting and analyzing personal behavioral data, such as internet activities, health information, or purchase patterns. It takes skill to strike a balance between gathering important informations and protecting people's privacy. To resolve these moral conundrums, it is essential to provide data anonymization, informed permission, and strong security measures. [15] [16].
- **Data Quality and Variability:** The caliber and variety of the data utilized for analysis have a significant impact on how well BA and AD perform. The accuracy and

dependability of AD systems might be hampered by noisy or insufficient datasets. Human behavior is also fundamentally complicated and vulnerable to changes that are impacted by environmental, social, and cultural influences. This variation can make it difficult to define common baseline behavior patterns and precise anomaly thresholds [15] [16].

- **Contextual Understanding and Interpretability:** Having a thorough awareness of context is essential for unraveling the complexities of human behavior. In some circumstances, behavior that may seem abnormal in one person may be completely normal on a nother. Accurate anomaly interpretation and contextual understanding require sophisticated algorithms that may change their workflow depending on the situation. Gaining trust and promoting efficient decision-making depend on the AD results being understandable and actionable [17].
- **Continuous Learning and Adaptation:** Because human behavior is dynamic and changes over time, BA and AD require ongoing learning and adaptation. To remain effective, models and algorithms must adapt to new behaviors and pattern-forming phenomena. The constant goal is to create systems that can learn from real-time data streams, change baseline behaviors without causing disturbances, and adapt to changing circumstances [17].

1.5 Computer Vision and Surveillance Intersection

As mentioned previously, CV plays a major role in security technology, especially in the surveillance task, where every aspect is directly related to controlling, analyzing, processing, and extracting useful information from video feeds or a sequence of images. Figure 1.7 is a description that illustrates the relation and the intersection between various disciplines to provide a powerful technology that gives the best performances and results [10]. Starting with the main domain of interests, the CV, which provides the algorithms and methods to do the main tasks related to security manners (OD, OT, activity recognition, and FR), On the other hand, ML techniques, tools, and algorithms give these tasks the power to be adaptable and can work with high efficiency in complicated scenes and scenarios, from learning to useful information extraction [10] [15]. As another part of the intersection, we have BA, which is a step to accomplish the surveillance task that analyzes movements and actions inside the scene, trying to understand and extract any anomalies or abnormal behaviors in order to take the right security measures. Finally, we have data sensors, which are the sources to provide the needed data to ML and CV algorithms and DL models in form videos, images, and audio data.

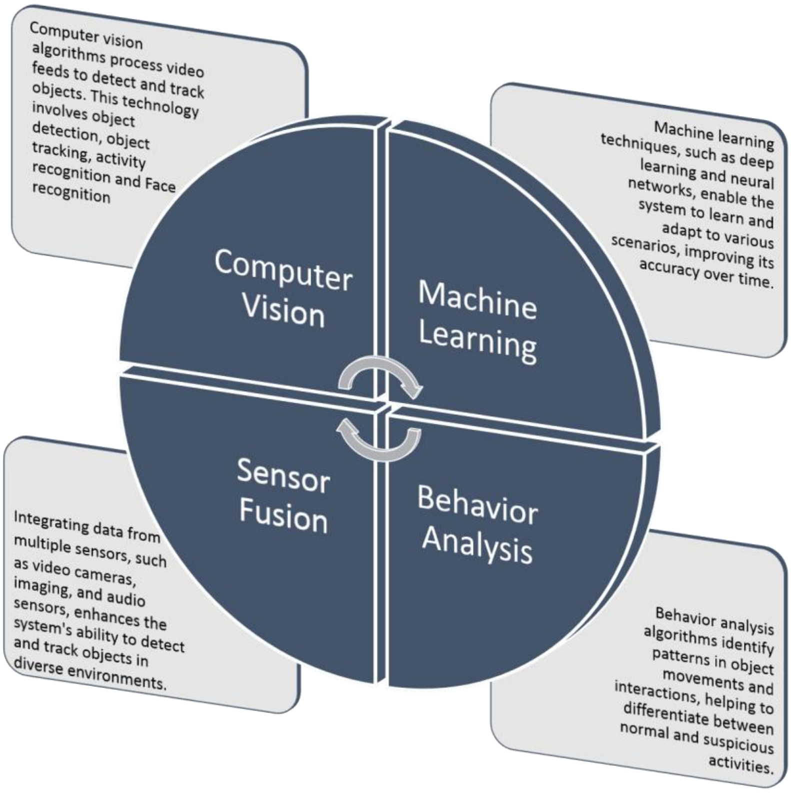

Computer vision algorithms process video feeds to detect and track objects. This technology involves object detection, object tracking, activity recognition and Face recognition
Machine learning techniques, such as deep learning and neural networks, enable the system to learn and adapt to various scenarios, improving its accuracy over time.
Computer Vision
Machine Learning
Sensor Fusion
Behavior Analysis
Integrating data from multiple sensors, such as video cameras, imaging, and audio sensors, enhances the system's ability to detect and track objects in diverse environments.
Behavior analysis algorithms identify patterns in object movements and interactions, helping to differentiate between normal and suspicious activities.

Fig. 1.7: The intersection between CV tasks and surveillance.

1.6 Smart City an Introduction

More than half of the world's population already lives in cities, and as technology continues to advance at an unprecedented rate, the idea of SC has emerged as a ray of hope for sustainable growth and improved quality of life in a fast-urbanizing world. SC is reshaping urban environments by enhancing efficiency, sustainability, and human well-being and livability through the integration of technology, data, and urban planning. Urbanization's quickening speed and technology's continuous advancement have drastically altered the world's terrain [18][19]. Today, more than 50% of people live in urban regions, which represents a significant demographic trend that is transforming communities all over the world. In this setting, the idea of a SC has emerged as a source of hope, providing a viable answer for sustainable urban development and an improved quality of life at a time marked by urban growth. In order to maximize efficiency and raise the general well-being and livability standards of its residents, a SC must be able to integrate ET, massive data resources, and meticulous urban design tactics. SCs are a practical reality shaping the cities of today and tomorrow by utilizing the power of interconnected systems and creative solutions [18].

The escalating problems caused by urbanization are inextricably related to the development of SCs. Cities are feeling the strain as more people move there in pursuit of better employment possibilities and living conditions. Providing an effective solution to these problems is what SC hopes to do through the thoughtful integration of technology. These cities may optimize resource allocation, streamline transportation networks, and improve public services by deploying a wide range of innovations, including IoT devices, AI-driven systems, and big data analytics. This complex web of technological developments promotes a setting where the urban environment is not only attentive to the demands of its inhabitants but also foresees and adjusts to them in real-time [18][19].

1.6.1 Smart City Defintion

SC is an application of IoT [18]. Urban living has undergone a radical transformation thanks to SC. These cities' fundamental strategy is to use data analytics and digital technologies to improve all facets of urban life, including infrastructure, transportation, energy use, healthcare, and public services. IoT, AI, data analytics, and sustainable practices are combined to build an ecosystem that improves the lives of citizens while reducing environmental effects [18][19]. SC aims to increase the general quality of life for its citizens while streamlining services and optimizing resources. In order to learn more about traffic flow, energy consumption, trash management, public safety, and other topics, this includes gathering and analyzing enormous volumes of data from numerous sources, including sensors, cameras, and other devices. City authorities may make wise decisions to solve urgent concerns and improve the sustainability, efficiency, and livability of their community by utilizing these data-driven insights [18][19]. Figure 1.8 represents what SC is about based on what IoT can provide from an essential element, which is data generation, data management, and application handling [18].

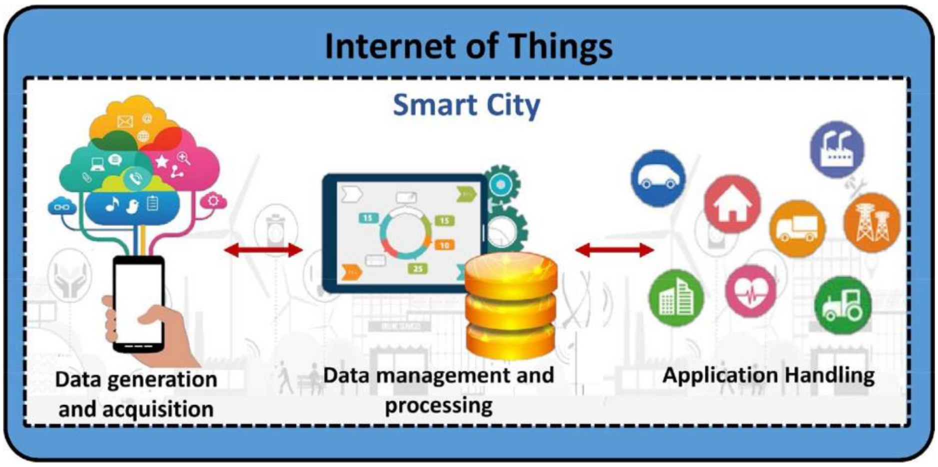

Internet of Things
Smart City
Data generation and acquisition
Data management and processing
Application Handling

Fig. 1.8: The relationship between IoT and SC [33].

1.6.2 Smart City Key Components

SCs thrive on the integration of the bellow listed key components, creating efficient, sustainable, and livable urban environments. By embracing technology, data-driven decision-making, and citizen engagement, SCs pave the way for a more connected and harmonious future for their residents, through:

- **IoT Devices:** A network of linked sensors, cameras, and other gadgets that collect information on a variety of municipal functions, including traffic, air quality, noise levels, and energy use [20].
- **Data analytics:** Sophisticated programs and algorithms that analyze the gathered data to find patterns, trends, and anomalies so that city authorities may make data-driven choices [20].
- **Cloud computing:** A centralized repository of data and applications that allows for safe access and teamwork amongst many stakeholders and departments [20][21].
- **Artificial intelligence:** ML systems and DL models that examine data to make predictions about the future, automate procedures, and provide citizens with individualized advice.
- **Open data platforms:** Websites that provide accessible datasets to promote openness, accountability, and cooperative problem-solving among enterprises, civil organizations, and government agencies [21]
- **Cybersecurity:** Strict security measures to safeguard confidential data and fend off online dangers while preserving the privacy and secrecy of personal information [20]
- **Collaborative governance:** Collaborations between local governments, businesses, academic institutions, and organizations of civil society to develop and carry out SC programs [20][21].
- **Citizen engagement:** Enable citizens to take part in decision-making, report problems, and offer suggestions for improving city services using Interactive technologies and platforms [21].

Figure 1.9 [20] summarizes those mentioned components, representing their main global domains that they belong to: IoT, Internet of Services (IoS), Internet of Data (IoD), and Internet of People (IoP). Table 1.1 [22][23] is a description of each one of the fields that the SC is about.

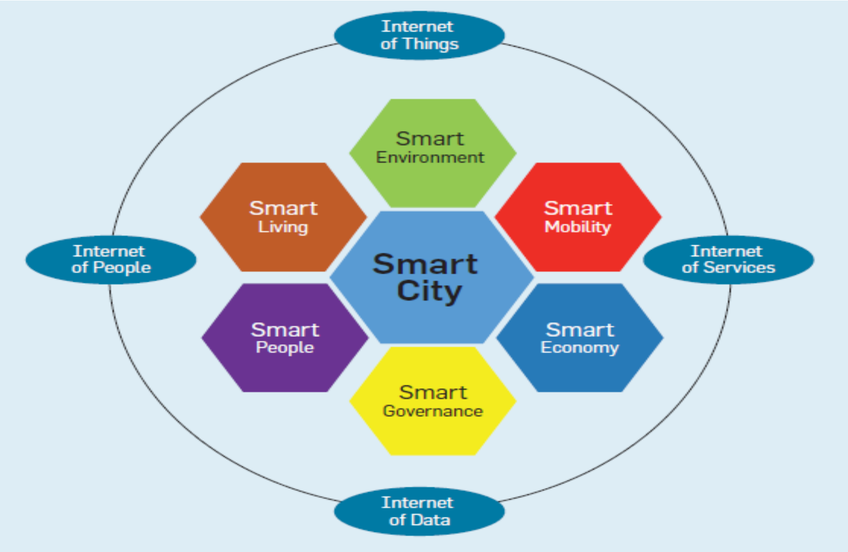


Fig. 1.9: SC Model [32].

Table. 1.1: Internet of things, Internet of services, Internet of data and internet of people.

IoT Every physical object that might be equipped with an IPv6 address.	▪ Communication objects based on internet technologies. ▪ Detection, identification, and location of physical objects ▪ Communication through connectivity.
IoS Service-based added value processes.	▪ New approach to providing internet-based services. ▪ Concepts for product-specific services on demand, knowledge provision, and services for controlling product behaviour. ▪ Interaction between people, machines, and systems to improve services.
IoD Manage big data integrate product and production data.	▪ Data is managed and shared using internet technologies. ▪ Cyberphysical systems are producing big data from the mountain, requiring the development of a holistic security and safety culture and establishing sustainable, trusted environments.
IoP Humans and their personal devices are active elements of applications, services, and network functions.	▪ **IoP is human-centric:** focusing on human social behavior (sociology, anthropology, cognitive psychology, etc.). ▪ **IoP is device-centric;** users' devices are seen as ''core IoP nodes'' (the devices and gadgets used by the users). ▪ **IoP is data- and computing-oriented:** which is what humans will use the Internet for.

1.6.3 Smart City Key Features

Thanks to recent advancements in digital technologies, SCs are now even smarter than before. SC has many technology components that are used by various applications in various sectors and many domains [21]. Figure 1.10 provides a summary of where the SC can be found, providing a step further in using the technological advances in our lives. Mentioning another main feature of SC among a lot of other points, such as the following:

- **Connected infrastructure:** Systems like SCs transportation networks, utilities, and public services can be monitored and managed in real-time thanks to interconnected infrastructure. This communication facilitates effective resource management and prompt problem-solving [22][24].
- **Data-driven insights:** The gathering and analysis of data is at the core of SCs. Large volumes of data are gathered by sensors, cameras, and other devices, which are then processed to produce insights that guide decision-making [21][22].
- **Sustainability and energy efficiency:** By incorporating renewable energy sources, maximizing energy usage, and lowering waste production, SCs place a priority on sustainability. This improves the quality of life for locals while reducing their carbon footprint [19][22].
- **Improved mobility:** SCs are characterized by effective transportation infrastructure. Congestion, pollution, and travel times are reduced via intelligent traffic management, real-time public transport tracking, and shared mobility services [22].
- **Citizen engagement:** Technology closes the participation gap between the general public and local governments. Citizens can express their concerns, report problems, and work together on community initiatives via digital channels [22].

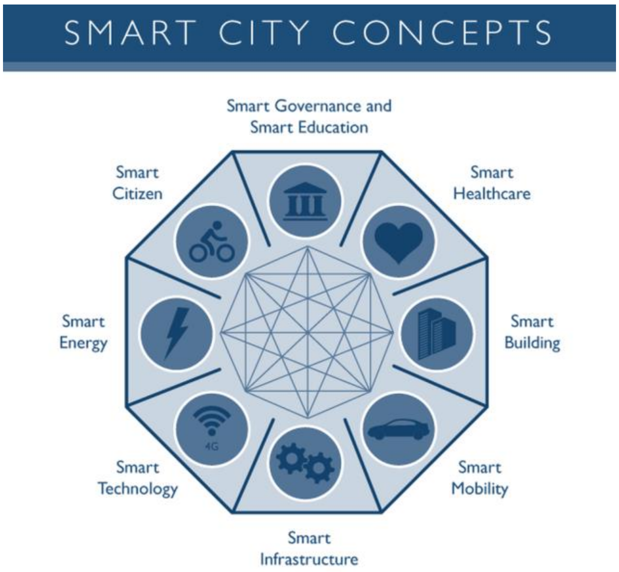


Fig. 1.10: SC Key Concepts[33].

1.6.4 Smart City Challenges and Considerations

SC development is a continuing process. Urban areas have endless possibilities for innovation as long as technology keeps advancing. Governments, corporations, and individuals must work closely together to fully realize the potential of SCs [20]. This entails putting in place precise legislation, fostering open data exchange, and raising public understanding of the advantages and consequences of SC programs. Since then, these cities are transforming how we live in and interact with our urban settings by leveraging the power of technology, data, and sustainable practices. The promise of increased productivity, sustainability, and citizen well-being drives cities' transition into smarter, more connected, and vibrant communities despite hurdles [20][25]. Nevertheless, some important points must be mentioned and taken into consideration:

- **Privacy and data security:** Concerns regarding citizen privacy and possible data misuse are raised by SCs significant data collection. Transparency and strong data protection safeguards are needed [25][26].
- **Inequalities in social access to technology**, or the "digital divide," could exacerbate existing ones. To close the digital divide and promote inclusivity, efforts must be made [20][25].
- **Interoperability:** Compatibility issues might arise when integrating distinct technologies and systems from various vendors. The maintenance of seamless interactions depends on standards and conventions [25][26].
- **Initial Investment:** Significant up-front expenditures in infrastructure and technology are necessary for the transformation to a SC. Although these expenditures are frequently outweighed by long-term advantages and financial planning [26].
- **Large-scale citizen data collection** and storage raises questions about privacy, surveillance, and the exploitation of personal data [26].
- **Security risks**: Include the possibility that hackers could compromise vital infrastructure and citizen data because of the blending of numerous systems and devices [26][27].
- **Ethical Considerations:** Careful consideration of ethical principles and values is necessary to ensure justice, accountability, and openness in algorithmic decision-making processes. [24][26].

1.7 Smart City Security

Security for SCs is a larger issue since, in order to provide it to the entire city, we must provide solutions for its particular applications, such as smart transport, smart health, smart environment, smart living, etc. An efficient, all-encompassing security solution for a SC will shield and defend the city from outside intruders while also securing its internal infrastructure, connectivity, applications, and services [28]. To address serious risks to security and privacy in smart environments, a variety of security and privacy solutions have been used, as detailed below.

1.7.1 Blockchain

The peer-to-peer distributed, centralized, and open decentralized blockchain technology is used to manage transactions such as purchases, sales, agreements, and contracts among many computers. A blockchain, in a nutshell, is a series of blocks where digital information is stored in a public database. It was designed particularly for cryptocurrencies like Bitcoin [29][30]. The security issues that arise with the creation of SCs have a compelling solution provided by blockchain technology. Due to its decentralized and unchangeable nature, it can improve data integrity, protect transactions, and enable ET applications that are advantageous to both residents and city officials. As SCs develop, utilizing the power of blockchain might hold the answer to improving urban environments for everyone's safety and security [29].

In the below-mentioned points, we list out some of the blockchain use cases in SC security [20][28][29]:

- **Supply Chain Management:** By tracking products from manufacture to distribution, blockchain can guarantee their validity. To protect consumers from fake or tainted goods, this is especially crucial for food, medicine, and other essential supplies.
- **Identity management:** Blockchain can help citizens create a secure, unchangeable digital identity. This can aid in guarding against identity theft and ensuring safe service access.
- **Decentralized data exchange** among cars, buildings, and traffic control systems can improve real-time traffic management, easing congestion and enhancing safety.
- **Energy system security:** Blockchain can improve the overall stability of the energy system, secure energy transactions, and allow peer-to-peer energy trade.
- **Emergency Response:** Blockchain can help emergency services and residents communicate securely and openly, ensuring that information is shared promptly and accurately.

1.7.2 Cryptography

Applications that are information-centric rely on algorithms that use cryptographic techniques to secure users' privacy and prevent illegal access to data while it is being stored, sent, and processed. Traditional encryption algorithms and standards are not fully appropriate for resource-constrained smart devices due to their energy consumption and computational complexity. Therefore, it is a basic requirement to use lightweight encryption, lightweight authentication protocols, etc. [20][29]. The importance of cryptography in guaranteeing security in the developing world of SCs, where data flows seamlessly and interconnection is the norm, cannot be overemphasized. Within the ecosystems of SCs, cryptographic approaches offer a solid foundation for the safety of sensitive data, preserving data integrity, and facilitating secure communication [29].

Some of the use cases of crptography in SC, among a lot, are listed below [20][28] [29]:

- **IoT security:** Cryptography protects data sent between IoT devices and centralized systems, ensuring that it is private and impervious to manipulation, preventing unauthorized access and control.
- **Smart grid security:** By limiting unwanted access to vital infrastructure, cryptographic solutions enable grid management, secure energy transactions, and protect the integrity of data on energy use.
- **Public safety and emergency response:** During emergencies, communication between emergency services and the general public can be done securely thanks to cryptography.
- **Traffic management:** Unauthorized manipulation of traffic flow is prevented via encrypted communication between vehicles, traffic lights, and central traffic management systems, which lowers accidents and congestion.
- **Collaboration and data sharing:** Cryptography enables safe data exchange between various city departments, fostering collaboration while preserving data privacy.

1.7.3 Biometrics

Authentication methods based on the IoT frequently employ biometrics. This technology heavily relies on human behavior and uses biometric information from faces, fingerprints, handwritten signatures, voices, etc. to automatically identify a person [20] [29]. Biometrics technology, which records and examines distinctive physical or behavioral characteristics, offers a potent answer. Using characteristics that are nearly impossible to duplicate in SC security, can be approached in a variety of ways using biometrics. SCs may use the potential of this technology to build urban settings that are not just effective and connected but also safe and secure for all people by resolving the issues related to biometric data usage and privacy concerns [29].

We are seeing a lot of smart applications based on biometrics, but some of the main industries are mentioned below [20][28][29]:

- **Smart access systems:** By using biometric technology to verify that only authorized individuals are allowed admission, in buildings, cars, and public areas, can lower the risk of unauthorized access and security breaches.
- **Healthcare services:** Biometrics can improve the security of medical information and patient identity verification, ensuring proper care and reducing medical identity theft.
- **Transportation security:** To improve security and passenger flow, biometric systems can be included in transportation infrastructures like airports and public transportation for identity checks, passport checks, etc.
- **Citizen services:** By decreasing paperwork and improving the citizen experience, biometrics can enable secure and effective access to government services.
- **Emergency response:** During disasters, prompt and precise victim and survivor identification can help emergency response activities, enhancing coordination and support.

When talking about the biometric security aspect, we are directly talking about the surveillance systems and the CV approaches to be applied as SC security. So, the next part will deal with both.

1.8 Surveillance in Smart Cities

The use of IoT technologies, which allow for the capture and analysis of massive volumes of data for immediate decision-making, is essential to the idea of SCs. However, there are substantial security, privacy, and surveillance issues that are brought up by the incorporation of IoT technologies into urban infrastructure [31]. In the context of SCs, surveillance refers to a wide variety of operations, including traffic monitoring, environmental sensing, public safety, and urban planning. An unprecedented amount of data is produced by the widespread use of IoT devices, including cameras, sensors, and wearable technology. This data can be used for a variety of purposes, such as the usage of surveillance cameras, which has become more common. While these cameras can offer useful information about human safety and public safety [24], This aspect of SC surveillance is represented and can be found in various domains and categories, mentioning the key ones as follows [20]:

- **Urban infrastructure management**: IoT-enabled sensors keep an eye on and control infrastructure assets like roads, bridges, and utilities in real-time, improving maintenance effectiveness and decreasing downtime.
- **Public safety and security**: Technology for gunshot detection, FR, and surveillance cameras all help with disaster management, emergency response, and criminal prevention.
- **Traffic management**: By providing real-time data on road conditions, parking availability, and public transportation, IoT devices help to optimize traffic flow, reduce congestion, and improve transportation systems.
- **Environmental monitoring**: Sensors keep tabs on waste management, noise levels, and air quality to help reduce pollution and conserve resources.
- **Healthcare and wellbeing**: Smart health monitoring systems and wearable gadgets enable people to track their health indicators and give medical experts useful information for individualized care.
- **Social services**: Using data analysis, local governments may distribute resources effectively and quickly address social issues.

Surveillance in SCs presents both opportunities and challenges. While it provides benefits that contribute to urban sustainability, safety, and quality of life. It also raises several critical challenges and civil liberties concerns that must be addressed to ensure individual rights and privacy [20][21]. Some of both points are mentioned below.

1.8.1 Benefits of Surveillance in Smart Cities

Surveillance in SCs offers numerous advantages, ranging from improved safety and efficient traffic management to better urban planning and resource utilization. When implemented responsibly and ethically, surveillance technologies contribute significantly to the creation of safer, more sustainable, and more efficient urban environments [18][20][24], susch as:

- **Effective resource management:** The gathering and analysis of real-time data optimizes the distribution of resources, reducing waste and raising overall effectiveness.
- **Crime prevention and public safety:** Surveillance tools help law enforcement authorities stop and deal with illegal activity, which ultimately improves public safety.
- **Sustainability of the environment:** IoT-enabled environmental monitoring encourages informed decision-making, which aids in the reduction of pollution and sustainable urban design.
- **Traffic Optimization:** By reducing traffic, travel times, and fuel consumption, intelligent traffic control systems increase urban mobility.
- **Rapid and effective emergency response** is aided by real-time surveillance data, saving lives in the event of accidents, natural disasters, and other emergencies.

1.8.2 Challenges and Concerns of Surveillance in Smart Cities

The ethical, technical, and social difficulties posed by IoT technologies necessitate a multidisciplinary approach. It is possible to achieve a balance between technology innovation and individual privacy by incorporating strong encryption techniques, resolving ethical concerns, minimizing monitoring creep, improving cybersecurity standards, and encouraging community interaction. However, this interconnected landscape between systems and devices has given rise to a myriad of concerns, particularly concerning privacy and security, mentioning the following [18][20][24]:

- **Privacy and data security:** Because of the risks associated with the widespread acquisition of personal data, encryption mechanisms and strong data protection measures are required.
- **Ethical considerations:** The use of biometric and FR technologies brings up moral issues with consent, profiling, and potential discrimination.
- **Monitoring creep:** Unintentional invasions of privacy may result from the gradual increase of monitoring capabilities without sufficient control.
- **Data breach:** Because IoT devices are interconnected, they are more susceptible to hackers, which could expose sensitive data.
- **Community engagement:** Forging trust and averting backlash from the public requires ensuring that citizens are involved in the decision-making process about surveillance efforts.

1.8.3 Computer Vision in Surveillance of Smart Cities

Computers can interpret and comprehend visual data using a variety of CV technologies. Automating human supervision is the main objective of CV applications for the security and surveillance industries. Modern surveillance systems in SCs are built on CV. These systems enable the automatic study and interpretation of visual data by utilizing techniques from AI, image processing, and ML. The CV function in surveillance is expected to transform urban living and radically affect how cities are controlled and experienced as SCs continue to develop [30]. A lot of applications and systems have been built as smart security solutions in different fields, domains, and in everyday life. Some of the main fields are listed below :

- **Perimeter monitoring and person detection:**

In SCs, person detection has a wide range of uses. Applications include real-time video analysis using CV techniques to decipher human activities. In restricted areas of airports or train stations for example, DL and CV algorithms are used to locate the target and detect intrusion events in real-time through the perimeter. For safety and security purposes, such automated, AI-based perimeter monitoring systems may effectively cover enormous monitoring regions [20][29].

- **Detect violent and dangerous situations:**

CV is frequently used in SC applications to automatically identify harmful situations, with the aim of protecting citizens through advanced video surveillance. Fighting, brawls, robbery, and other potentially risky situations. A difficult issue with these scenes is their high unpredictability. In order to recognize human behavior and interactions, algorithms are utilized for people detection, tracking, and three-dimensional human position estimation. In order to describe scenarios and events that should be triggered to automate human intervention or provide analytics, the output of AI models is processed via a logic flow. These applications include people counting and dwell time analysis, which is the detection of people who spend an extraordinary amount of time in particular locations [20][21][29].

- **Protection of critical infrastructure in smart cities :**

Intelligent computing technologies are used in SCs to keep an eye on activity or head off dangers. Society's critical infrastructures are resources that are absolutely necessary, and their collapse would be extremely costly and disruptive. Roads, communications, water, energy, and other things are some of them. While closed-circuit television systems are now a crucial component of security and law enforcement, conventional surveillance systems rely on a human operator's ability to pay attention when faced with numerous video streams. The main drawbacks of conventional video surveillance systems include high running expenses because staff must examine the footage, mistakes brought on by fatigue, and a massive volume of video data requiring large bandwidths. In order to develop large-scale systems, new edge AI technologies enable distributed sensing systems to achieve real-time performance, cost-efficiency, and scalability. Running ML on-device with nearby edge nodes that connect to the cloud is made possible by edge AI [20][21].

- **COVID-19 Pandemic and Smart City:**

The coronavirus pandemic has highlighted the shortcomings of current urban planning. Therefore, ET methods and technology like CV offer quick and practical ways to mitigate pandemic dangers and a lot for providing more safety to people's lives. We mentioned some other smart solutions, such as:

- **Crowd disaster avoidance application:**

Due to the rise in public meetings where there is a high risk of disasters and stampedes, CV monitoring in crowded settings is one of the most crucial applications. In order to increase public safety, vision-based crowd catastrophe prevention technologies are deployed. These applications typically concentrate on crowd scenes and BA. Stampedes can happen as a result of unforeseen circumstances or people acting abnormally. With one or more cameras at scale, we have porous models thanks to DL models, which are utilized to count people and estimate crowd density [32][33].

- **Social distancing monitoring**

In order to effectively restrict the spread of the epidemic, smart vision technologies are capable of observing and enforcing social distance between individuals. For instance, applications for social distancing monitoring with real-time OD to recognize humans in videos of surveillance cameras at scale can be made using DL-based AI models for crowd recognition can automatically find people in urban public settings. The movement patterns of people are utilized to detect potentially high-risk areas so that open and public spaces might be redesigned to be more pandemic-proof [32][33].

- **Automated mask detection**

Automated mask detection of people in public spaces is another example of the applications. In addition to complying with social distancing rules in public locations like subway stations, AI-enabled systems can enhance mass surveillance by utilizing DL algorithms to monitor the conditions and give alarms if people do not wear a mask or do not comply with lock-down measures. The information can be used to improve cities' ability to anticipate pandemic patterns, enable prompt action, reduce virus spread, support certain industries, lessen disruption of the supply chain, and guarantee the continuation of vital operations and services[33][34].

1.9 Conclusion

This chapter has delved into two main concepts of CV and SC. Starting with the definition of the CV domain, its relationship with the surveillance security manner, and showing the main problems faced in this field from OD, OT, BA, and AD, where we represent for each its proper definition, how it works, and examples of applications, to show later the intersection between CV and surveillance. As the second point of this chapter, we cover the SC concept from definition to key components, features, and challenges. Also, we cover the relation between this domain and the security aspect in general and the security aspect in particular. Overall, the insights gained in this chapter form a crucial link in the chain of knowledge construction and bring us one step closer to addressing the broader research objectives outlined in this thesis.

Chapter 02 : Intelligent Surveillance Systems

2.1 Introduction

In an age where security is vital, technology continues to play a pivotal role in expanding our abilities to monitor and safeguard our surroundings. Ensuring security and safety is a top priority for individuals, businesses, and governments alike. Traditional surveillance systems have come a long way, but the demand for more advanced and efficient solutions has never been greater. The integration of IoT and ET into surveillance systems has led to the development of intelligent surveillance systems (ISS) that are more intelligent, efficient, and diverse than ever before. These solutions are empowered by IoT, ET, and AI to create systems that are revolutionizing the way we monitor and protect our surroundings by offering real-time insights, enhanced automation, and reduced response times [35].

2.1.1 Edge Technology

The idea behind ET, also called edge computing, is to gather, store, process, and analyze data closer to the point of use in order to speed up reaction times and conserve bandwidth. ET is a distributed computing system that enables applications to be run closer to data sources such as edge servers, IoT devices, and local endpoints. Figure 2.1, represented below, is an illustration of ET and how this infrastructure is conducted.

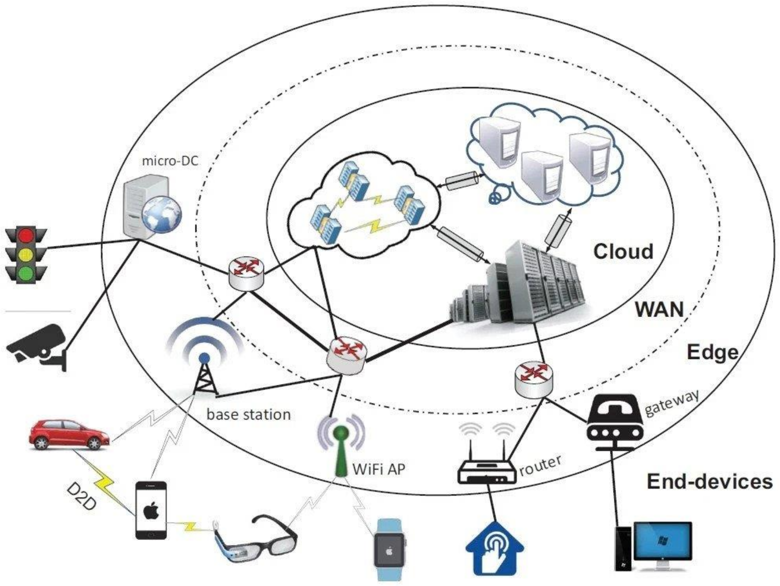


Fig. 2.1: An illustration of ET [27].

It reflects the basic principle of ET, which moves cloud services from the network core to the network edges that are in closer proximity to IoT devices and data sources. An edge node can be a nearby end-device connectable through device-to-device communications, a server attached to an access point (e.g., WiFi, router, and base station), a network gateway, or even a microdatacenter ready for use by adjacent devices. While edge nodes can vary in size, ranging from a credit card, smart watch, smart cars, smart phones, etc [35].

Talking about ET leads us to also talk about cloud computing and the difference between those two. ET and cloud computing are not mutually exclusive. Instead, the edge complements and extends the cloud. The key advantages of merging ET with cloud computing are the following [36]:

- **Backbone network performance**: Distributed ET nodes may handle numerous calculation jobs without exchanging the underlying data with the cloud. This provides for optimizing the traffic load of the network.
- **Services real-time response**: Intelligent applications deployed at the edge can drastically reduce the delay of data transmissions and improve the response speed.
- **Powerful cloud backup**: In circumstances where the edge cannot afford it, the cloud can provide powerful processing capabilitie and scalable storage. Data is increasingly created at the edge of the network, and it would be more efficient to also process it the data at this edge.

Hence, ET is a key option to break the bottleneck of new technologies based on its advantages of lowering data transmission, improving service latency, and alleviating cloud computing pressure.

2.1.2 Edge Intelligence and Edge Artificial Intelligence

The merger of ET and AI has given rise to a new study topic termed “Edge Intelligence” or “Edge AI”. Edge AI makes use of broad-edge resources to enable AI applications without fully relying on the cloud. While the phrase edge AI or edge intelligence is brand new, activities in this area have begun early in this century. However, despite the early commencement of the investigation, there is still no precise definition of edge intelligence. Currently, most organizations and media refer to edge intelligence as “the paradigm of running AI algorithms locally on an end device with data (sensor data or signals) that are created on the device” [37]. Edge AI is well-acknowledged topics for research and commercial innovation. Due to the superiority and necessity of executing AI applications on the edge, it has recently gained tremendous interest. Gartner Hype Cycles describes edge intelligence as an emerging technology that will achieve a plateau of productivity in the following 5 to 10 years [38]. Multiple large organizations and technology leaders, including Google, Microsoft, IBM, and Intel, highlighted the advantages of ET in bridging the final mile of AI. These efforts cover a wide range of AI applications, such as real-time video analytics, cognitive aid, precision agriculture, SC, smart home, and industrial IoT [37].

2.1.3 Artificial Intelligence of Things

Artificial Intelligence of Things (AIoT) combines the connectivity of the IoT with the data-driven knowledge derived from AI. This new technology is centered on the integration of AI with IoT infrastructure. By merging these two, the data generated by dispersed nodes can be utilized by applying AI techniques such as ML algorithms and DL models. As a result, ML capabilities are shifted closer to the data source. We can say that this notion refers to edge AI, or edge intelligence [35].

Since the edge is closer to users than the cloud, ET is expected to solve numerous difficulties. In fact, ET is gradually being merged with AI, benefiting each other in terms of the realization of edge intelligence and intelligent edge, as indicated in Figure 2.2 [39]. Edge intelligence and intelligent edge are not independent of each other. Edge intelligence is the goal, and the DL services in intelligent edge are also part of edge intelligence. In turn, an intelligent edge can deliver improved service throughput and resource utilization for edge intelligence.

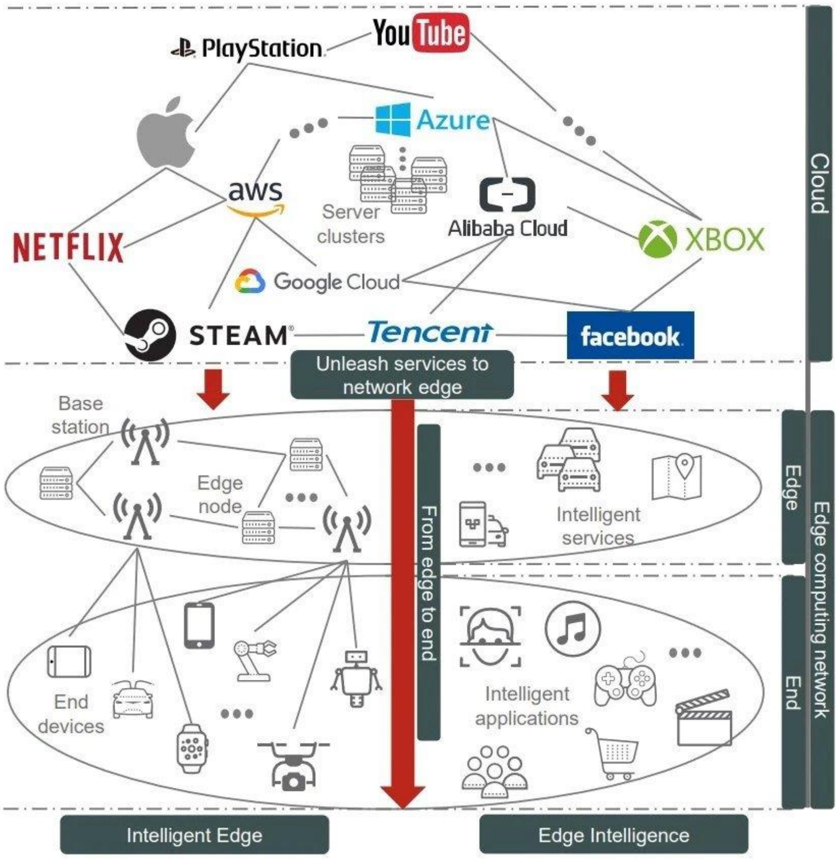


Fig. 2.2: Edge intelligence and intelligent edge [28].

These paradigms mentioned above are the key to the convergence of surveillance in various aspects, as follows:

- **Real-time data collection:** ISS is designed to capture real-time data from a network of interconnected devices. These gadgets include cameras, sensors, and other IoT-enabled components. This data is useful for monitoring and responding swiftly to security threats. For instance, cameras outfitted with IoT capabilities can collect high-resolution photos and videos, which are instantly transmitted for analysis [40].
- **ET for swift decision making:** One of the most significant breakthroughs in surveillance technology is ET. Instead of sending all data to a centralized server for processing, edge devices (such as cameras) can handle data locally. This means that key choices can be made in real-time without relying on a distant server. ET improves latency and minimizes the risk of data loss due to network outages [40].
- **Enhanced automation:** IoT devices in surveillance systems are equipped with AI algorithms that may recognize strange patterns, objects, or actions. For example, if an unauthorized individual enters a restricted area, the system can trigger alarms, notify security staff, or even perform automated responses like locking doors or turning on lights [39][40].
- **Connectivity beyond imagination:** IoT empowers surveillance by integrating an assortment of devices, such as cameras, sensors, and access control systems, into a cohesive network. These interconnected devices may interact fluidly with each other, providing a holistic view of the monitored region. This link extends to the internet, enabling remote monitoring and control from nearly anywhere in the world [40].

2.2 Intelligent Surveillance System Components

In an era where security and surveillance are vital, ISS have evolved as strong tools for securing people, property, and investments. These systems are not just about cameras gathering images; they are sophisticated networks of components working together to deliver real-time insights, automation, and increased protection. We will discuss the essential components that make up ISS, throwing light on how they function in harmony to create a safer and more efficient society [41].

- **Cameras with advanced sensors:** The heart of any surveillance system are the cameras. These cameras are fitted with modern sensors, including high-definition video cameras, infrared sensors for night vision, and motion detectors. Some cameras even have thermal imaging capabilities, which can detect heat signatures. These sensors enable the system to acquire visual data effectively [40].
- **Data storage and sanagement:** The huge volume of data collected by surveillance cameras requires sophisticated storage and management systems. Modern systems leverage network video recorders or cloud-based storage to store video footage. Additionally, efficient data management techniques help organize and retrieve data efficiently [40][41].
- **Network Infrastructure:** To send data from cameras to storage and monitoring devices, a robust network infrastructure is necessary. This covers wired and wireless

connections, such as Ethernet cables, Wi-Fi, and cellular networks. High-speed data transport provides real-time monitoring and speedy response [41].

- **Video analytics and AI algorithms:** The power of ISS rests in its capacity to evaluate data in real-time. Video analytics and AI algorithms process video feeds, detecting objects, people, and patterns. These systems can spot suspicious behavior, such as invasions or unauthorized access, and generate alarms or automatic replies [40][41].
- **Edge computing devices:** ET devices play a significant part in ISS. These devices are situated near the network's edge, close to the cameras. They may process data locally, lowering latency and minimizing the need for frequent contact with centralized servers. ET offers real-time decision-making, making surveillance systems more efficient [40].
- **Monitoring and control centers:** Human operators play a crucial part in surveillance systems. Monitoring and control centers are outfitted with monitors that provide live video feeds, analytics data, and alarms. Operators can respond to situations in real-time, summon security personnel, or conduct automated actions. [40][41].
- **Alerting and automation:** When the system identifies a potential threat or odd activity, it can activate alerts in numerous ways, such as sirens, email messages, or mobile app alerts. Automation capabilities let the system conduct specified actions, including locking doors, turning on lights, or tracking an intruder's movement [41].
- **Integration with other systems:** ISS often interfaces with other security and access control systems. This integration enables a cohesive security environment where cameras function in concert with access control, alarms, and visitor management systems to increase overall security [41].

ISS has gone a long way from simple security cameras. They are now full ecosystems of interconnected components meant to deliver increased security, real-time information, and automation. By integrating modern cameras, data analytics, ET, and solid network infrastructure, these systems offer unparalleled capabilities for safeguarding assets and guaranteeing public safety. As technology continues to improve, we can expect clever surveillance systems to become even more clever and useful in our increasingly complicated society.

2.3 IoT Micro-controllers: The Brain of Intelligent Surveillance Systems

In the area of ISS, where security and situational awareness are paramount, IoT microcontrollers play a pivotal role as the unsung heroes behind the scenes. These tiny yet powerful gadgets act as the brain of surveillance systems, providing networking, data processing, and real-time decision-making [42]. They provide a lot of other advantages and significance in the ISS, shining light on their capabilities and impact on modern security solutions. IoT microcontrollers are compact computing light weight devices that integrate processing power, memory, communication interfaces, and sensor inputs onto a single chip. They serve as the foundation upon which ISS are built, providing the following key functionalities [42][43]:

- **Data collection:** IoT microcontrollers are equipped with sensor interfaces that allow them to gather data from multiple sources. In surveillance systems, these sensors include cameras, motion detectors, temperature sensors, and more.
- **Data processing:** Once data is acquired, IoT microcontrollers are responsible for processing it in real-time. They can conduct activities such as image and video analysis, sensor data interpretation, and even advanced analytics using onboard processing capabilities or by interacting with other servers.
- **Connectivity:** Connectivity is at the foundation of IoT microcontrollers. They are meant to connect to the internet or local networks utilizing protocols like Wi-Fi, Ethernet, or cellular data. This connectivity enables remote monitoring, data transmission, and interaction with other devices and systems.
- **Operating at the edge:** One of the most significant advantages of IoT microcontrollers is their ability to perform ET. Allowing data processing to occur at the device itself, minimizing latency and dependence on centralized servers. In surveillance, this entails speedier decision-making and real-time notifications.
- **Control and automation:** IoT microcontrollers can execute specified actions based on the data they collect and process. For instance, when a surveillance camera detects motion, the microcontroller can trigger an alarm, activate other cameras, or send notifications to security personnel.
- **Energy efficiency:** Many IoT microcontrollers are designed for minimal power consumption, making them perfect for battery-powered surveillance equipment such as wireless cameras and sensors. This energy efficiency ensures continuous operation without frequent battery replacements.

There are various IoT microcontrollers appropriate for surveillance applications, each with its own unique set of features and capabilities. The choice of microcontroller depends on the specific requirements of your monitoring system. Here are some popular IoT microcontrollers often used in surveillance:

- **Raspberry Pi:**

Raspberry Pi [44] boards, such as the Raspberry Pi 4, are adaptable and commonly used for IoT-based surveillance projects. They feature a strong processor, enough RAM, and many networking choices, including Ethernet and Wi-Fi. Raspberry Pi can run a number of operating systems and support camera modules for video capture (Figure 2.3).

Fig. 2.3: Raspberry Pi 4 [44].

- **Arduino:**

Arduino boards [45], notably the Arduino MKR series, are noted for their simplicity and ease of use. While they may not have the computing power of the Raspberry Pi, they are sufficient for basic surveillance jobs. Arduino boards can link with sensors, cameras, and other components to collect and transfer data (Figure 2.4).

Fig. 2.4: Arduino UNO [45].

- **ESP32:**

The ESP32 [46] is a popular IoT microcontroller recognized for its Wi-Fi and Bluetooth features. It's excellent for IoT surveillance applications where wireless communication is vital. ESP32-based boards can record and transmit video and sensor data to a central server (Figure 2.5).

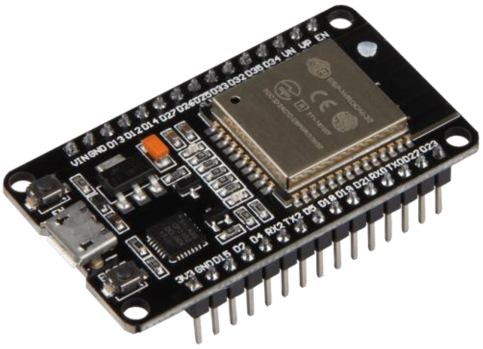


Fig. 2.5: ESP32 [46].

- **ESP8266:**

The ESP8266 [47] is a cost-effective Wi-Fi microcontroller used for simple surveillance applications, such as sensor data collection and transfer (Figure 2.6).

Fig. 2.6: ESP8266 [47].

- **NVIDIA Jetson Nano:**

The NVIDIA Jetson Nano [48] is a powerful microcontroller developed for AI and CV applications. It's perfect for advanced surveillance systems that demand real-time video analytics and OD. Jetson Nano boards provide GPU acceleration for AI tasks (Figure 2.7).

Fig. 2.7: NVIDIA Jetson Nano [48].

- **BeagleBone:**

BeagleBone [49] boards, like the BeagleBone Black, provide a blend of power and versatility for surveillance projects. They feature extensive I/O possibilities and can run Linux distributions, making them appropriate for camera integration and data processing (Figure 2.8).

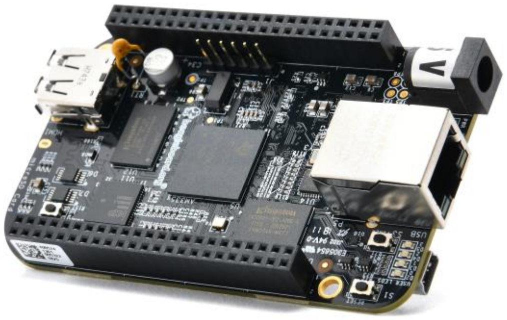


Fig .2.8: BeagleBone [49]

- **Particle Photon and Electron:**

Particle [50] offers IoT-focused microcontrollers with integrated cellular connectivity. These devices are suited for remote surveillance applications where an internet connection may not be easily available (Figure 2.9).

Fig. 2.9: Particle Photon and Electron [50].

- **Odroid:**

Odroid [51] boards, such as the Odroid XU4, are powerful single-board computers appropriate for surveillance activities. They include many USB ports and ethernet connectivity for camera integration and data transfer (Figure 2.10).

Fig. 2.10: Odroid [51].

- **Helium Atom:**

The Helium Atom [52] is a low-power, long-range IoT microcontroller appropriate for outdoor surveillance applications. It employs the Helium wireless network for communication (Figure 2.11).

- **Microchip PIC and AVR Series:**

Microchip's PIC and AVR [53] series microcontrollers are noted for their low power consumption and wide variety of sensor interface capabilities. They are suitable for battery-powered surveillance sensors and gadgets (Figure 2.12).

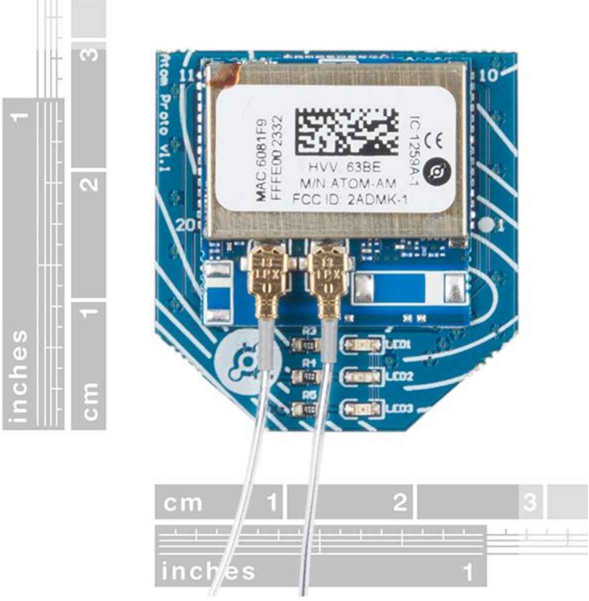


Fig. 2.11: Helium Atom [52].

Fig. 2.12: Microchip PIC and AVR [53].

When picking an IoT microcontroller for surveillance, we must take into consideration points such as processing power, networking options, power requirements, and compatibility with sensors and cameras. Additionally, the option may depend on whether your surveillance system requires simple monitoring or advanced features like video analytics and AI.

2.4 Real-world Applications of Intelligent Surveillance Systems

ISS is increasingly being deployed in various real-world scenarios and applications to enhance security, improve efficiency, and provide valuable insights. Here are some real-world examples of this smart systms [54]:

- **Smart home security systems:** Smart home security systems use IoT surveillance technology to protect houses. Examples include video doorbell cameras (e.g., Ring, Nest) that allow homeowners to watch and converse with visitors remotely, indoor/outdoor security cameras that give real-time video surveillance, and smart locks that can be operated remotely via a smartphone app (Figure 2.13).

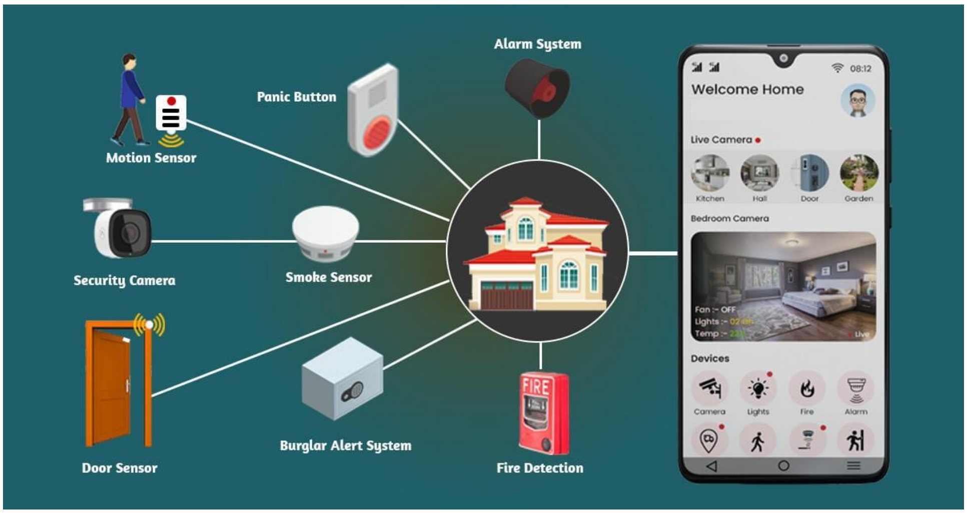


Fig. 2.13: Illustration of smart home surveillance applications [54].

- **Retail loss prevention:** Many retail businesses utilize smart surveillance systems to prevent theft and improve client safety. These systems generally include video analytics to detect suspect behavior, measure foot traffic patterns, and check inventory levels. In the event of a security breach, alarms can be triggered and alerts sent to security staff (Figure 2.14).

Fig. 2.14: Illustration of retail smart services and loss prevention [55] .

- **Traffic management and monitoring:** Smart surveillance is used to monitor traffic conditions in cities. Cameras integrated with license plate recognition technology help identify and track vehicles. Traffic signals can be modified in real-time based on traffic flow, and issues such as accidents or congestion can be identified and addressed swiftly (Figure 2.15).

- **Public transportation security:** Public transportation systems, such as subways and buses, use smart monitoring to enhance passenger safety and prevent vandalism. Surveillance cameras are installed in stations and vehicles to monitor operations, deter criminal behavior, and offer proof in case of incidents (Figure 2.16).

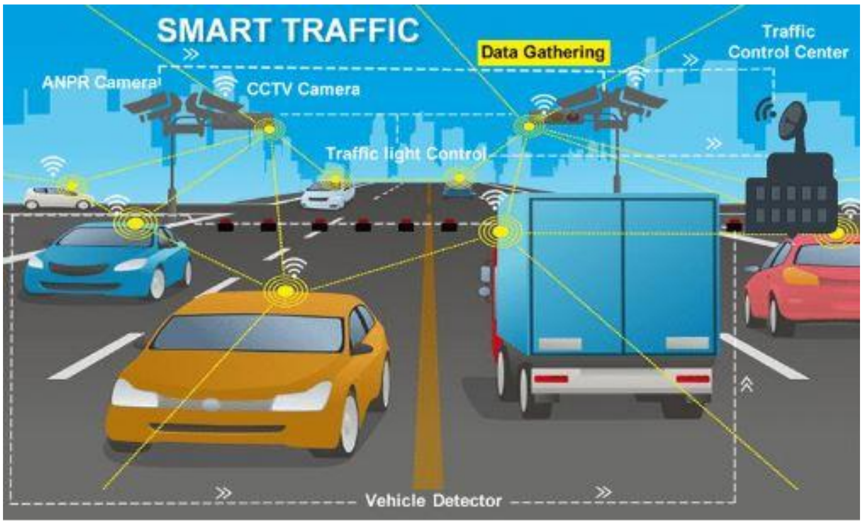


Fig. 2.15: Illustration of traffic management and monitoring [54].

Fig. 2.16: Illustration of public transportation smart security [55].

- **Healthcare facilities:** Hospitals and healthcare facilities utilize smart surveillance to increase patient safety and security. Cameras and sensors are deployed to monitor sensitive locations like the emergency room, ensuring fast reactions to emergencies and tracking the movement of patients and employees (Figure 2.17).

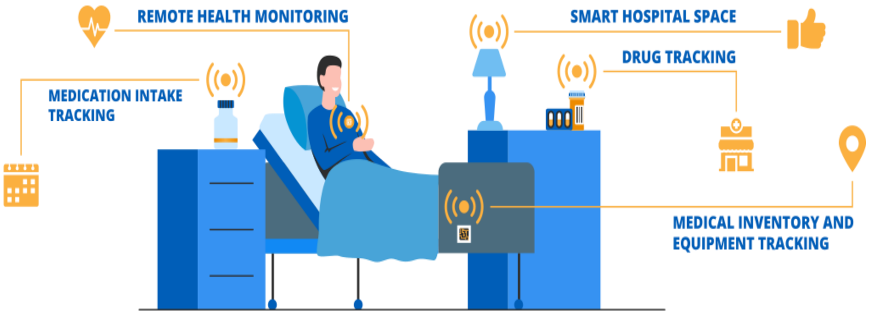

REMOTE HEALTH MONITORING
SMART HOSPITAL SPACE
DRUG TRACKING
MEDICATION INTAKE TRACKING
MEDICAL INVENTORY AND EQUIPMENT TRACKING

Fig. 2.17: Illustration of public health smart security [55] .

- **Agriculture and Farming:** In agriculture, ISS help monitor crops and cattle. Cameras and sensors can detect indicators of illness, pests, or water stress in crops. They can also provide insights into animal behavior and health (Figure 2.18).

Fig. 2.18: Illustration of agriculture and farming smart security [54].

- **Environmental monitoring:** Environmental authorities utilize smart surveillance for monitoring natural calamities, such as wildfires and floods. Drones equipped with cameras and sensors can provide real-time data for early warning systems and disaster management (Figure 2.19).
- **Airports and transportation hubs:** Airports and transportation hubs rely on smart monitoring for security and operational efficiency. FR technology, especially is used to identify and track passengers, boosting security and streamlining check-in and boarding processes (Figure 2.20).

Fig. 2.19: Illustration of environmental monitoring using smart surveillance [54].

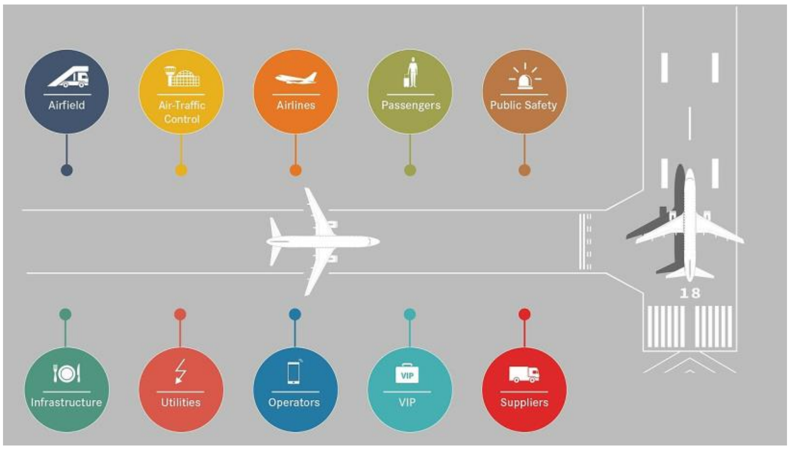


Fig. 2.20: Illustration of airports and transportation hubs with smart security[55].

2.5 Conclusion

An in-depth examination of the ISS field has been done in this vital chapter, with a special emphasis on the ET constituting a critical technological sector. Beginning with a specific definition of ET, the context for subsequent discussions is established by a thorough investigation of related concepts such as Edge Intelligence, Edge AI, and AIoT. Following that is a careful dissection of the ISS components, revealing the intricate workings of this equipment. Microcontrollers, the ISS's figurative brain, are at the heart of these systems. The chapter digs into the subtle features of these microcontrollers, providing a general understanding of their crucial role in propelling ISS. Numerous exemplars are offered, each accompanied by a definition, establishing a tangible link between theory and applications.

Chapter 03 : Facial Recognition for Smart Cities Surveillance

3.1 Introduction

FR technology, a subset of biometric technology and CV domain, is a field that has witnessed remarkable growth and transformation over the past two decades. From its humble beginnings as a novel concept in computer science to its widespread applications in today's digital world, FR has become an integral part of our daily lives [55]. FR, simply put, is the process of identifying or verifying individuals by analyzing and comparing their facial features. This technology uses computer algorithms to map and measure distinctive facial characteristics, such as the distance between the eyes, the shape of the nose, and the contour of the jawline. These facial features are then converted into a unique digital representation known as a face template, faceprint, or signature [56]. The fundamental goal of FR is to match these faceprints against a database of known faces to identify individuals or verify whether a person is who they claim to be.

The 21st century marked a significant turning point for FR. Several factors contributed to its rapid progress:

- **Data Availability and Benchmarks:** The proliferation of digital cameras, social media, and surveillance systems led to the collection of vast amounts of facial data. This abundance of data enabled researchers to train more accurate and robust FR algorithms. Benchmark availability, which provides a standardized dataset for researchers to compare the performance of different methods, also contributed to this progress [56].
- **Deep Learning:** DL algorithms have revolutionized the way computers perceive and process facial information. One of the primary contributions of DL to FR is its ability to automatically learn hierarchical features from raw image data. Traditional FR systems often relied on manually crafted features, which were limited in their ability to handle the complexity and variability of human faces. DL algorithms, such as CNNs, excel at automatically detecting and extracting intricate facial features, making them exceptionally effective for recognizing faces under various conditions. Moreover, DL models are highly adaptable. They can be trained on large datasets containing diverse facial images, enabling them to generalize and perform well across a wide range of facial variations, including different poses, expressions, and lighting conditions. This adaptability has made DL-based FR systems far more robust and versatile than their predecessors[56][57].
- **Hardware:** As the demand for accuracy, speed, and real-world applicability has grown, so too has the need for specialized hardware tailored to the unique requirements of FR systems. One of the key hardware breakthroughs that has greatly enhanced FR is the development of graphics processing units (GPUs) and specialized AI accelerators. These hardware components excel at handling the complex mathematical computations involved in DL models, which form the backbone of modern FR algorithms. GPUs, in particular, have significantly expedited the training and inference processes, making it possible to create highly accurate and responsive FR systems in real-time. Additionally, 3D sensors and cameras have expanded the capabilities of FR systems. 3D hardware captures depth

information, allowing for more robust recognition, even in challenging scenarios. This technology has been pivotal in making FR systems more secure and reliable [56][57].

- **Commercial Applications:** Companies like Apple and Facebook have integrated FR into their products, making it a mainstream technology. Apple's Face ID, introduced in 2017, brought FR to millions of consumers, highlighting its convenience and security [56].

This FR technology is widely used for security, authentication, and convenience purposes, and it has applications in various industries, including law enforcement, banking, and mobile device security. FR has come a long way from its early days as an experimental technology. It has become a ubiquitous part of our digital lives. However, it also raises important questions about privacy, ethics, and bias that continue to be debated and addressed. As we move forward, the evolution of FR will likely be shaped not only by technological breakthroughs but also by societal and ethical considerations [55][57].

Here is a concise overview of the history of FR from 1960 to 2020, highlighting key developments and milestones [55][56][57]:

- **1960s-1970s:** The nascent stage FR research began in the 1960s and 1970s with basic experiments in pattern recognition. Early efforts were limited by computational capabilities and lacked practical applications. The American researchers Bledsoe et al. [58] are considered to be the pioneers of FR, introducing at that time a system to classify face photos by recording the locations of facial features. In the 1970s, Harmon et al. [59] made the system more accurate by detecting faces automatically using biometrics recorded manually. They used 21 facial markers that included the thickness of the lip and the color of the hair.
- **1980s:** Researchers started exploring the potential of using computers to identify and analyze facial features, using AI that was introduced to develop previously used theoretical tools, which showed many weaknesses. Mathematics ("linear algebra") was used to interpret images differently and find a way to simplify and manipulate them independently as human markers.
- **1990s:** In the early 1990s, Matthew Turk and Alex Pentland introduced the Eigenface [60] method, which was a significant breakthrough. This method employed principal component analysis (PCA) to extract essential facial features, marking an important milestone in the field.
- **1998-2000s:** The Defense Advanced Research Projects Agency and the National Institute of Standards and Technology started the FR technology program (FERET) to encourage commercial use of FR technology. By 2002, law enforcement officials had deployed FR for critical testing.
- **2000-2010s:** The Fisherface [61] method in 2001 for FR was created; Viola-Jones' face detection framework [62] was introduced in 2004; The Face Recognition Grand

Challenge was launched in 2006, with a primary goal of promoting and advancing FR technology in the U.S. government; and also in the same year, Geoffrey Hinton and his team introduced the Restricted Boltzmann Machine [63], a building block for DL networks. The introduction of the Labeled Faces in the Wild (LFW) benchmark dataset in 2007 provided researchers with a standardized dataset for evaluating and comparing FR methods. The Deep Belief Network a probabilistic generative model, shows promise in feature learning and representation. Commercial applications of FR by Facebook in 2010 helping to identify people's IDs by uploading their photos.

- **2011-Current:** The FR technology takes a big step forward thanks to DL approach that works based on deep neural networks (DNN). AlexNet (2012) [64], a deep CNN, achieves a breakthrough in image classification tasks. Google's FaceNet (2014) [65] introduces the concept of learning a high-dimensional face embedding, enabling accurate FR using DNN. Facebook released the DeepFace (2014) [66] model, demonstrating remarkable accuracy in identifying faces inside images. VGGFace (2015) [67], based on the VGG architecture, achieves competitive performance in FR on various datasets. DeepID (2015) [68], developed by researchers at the Chinese University of Hong Kong, presents a novel approach to face verification using DL. ArcFace (2018) [69], a margin-based FR framework, advances the state-of-the-art in FR accuracy.
- **2020:** The year 2020 saw ongoing advancements in FR domain, including improvements in accuracy, 3D FR technologies [70], and a focus on recognizing faces even when individuals wear masks, in response to the COVID-19 pandemic. Researchers and policymakers have also increasingly addressed ethical and privacy concerns related to FR technology.

Figure 3.1 illustrates the timeline, highlighting how FR has evolved from its early experimental stages to becoming a ubiquitous technology with applications in various sectors, including security, mobile devices, and more.

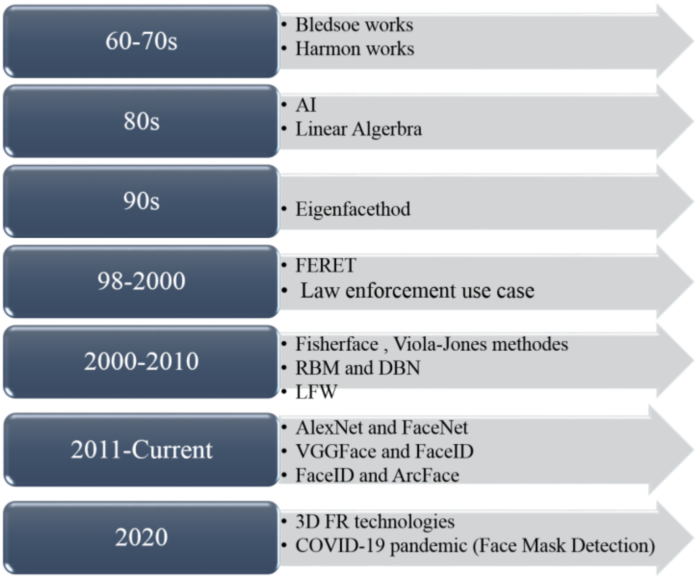


Fig. 3.1: FR timeline progress.

3.2 Face Recognition System

The functionality of the FR system entails the creation of additional subtasks in a pipeline that are essential for its proper execution [56][57]; an illustration of these subtasks is presented in Figure 3.2.

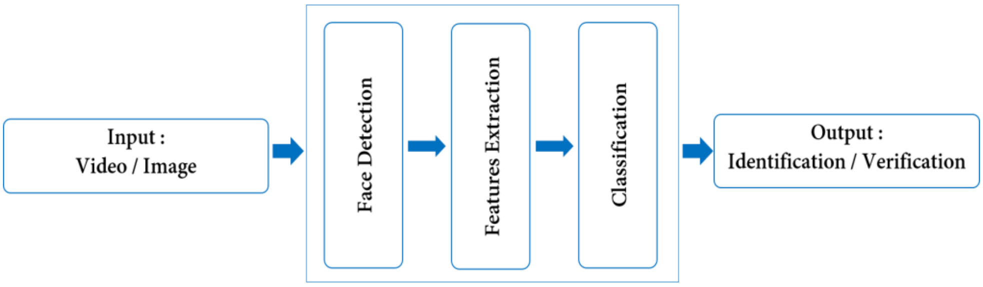


Fig. 3.2: The standard design of an automated FR system.

1. **Face detection:** It's the initial phase in the system; it shows if the image contains a face(s) or not by determining human facial location and dimensions. A lot of methods and algorithms are proposed for this task; some of them, among others, are the Haar-Cascade (HC) [62] method by Viola and Jones to identify facial characteristics with Haar-like features. Dlib Histogram of Oriented Gradients (HOG) [71], which employs support vector machines (SVM) in conjunction with histograms of oriented gradients. Dlib CNN [71] combines a maxi-mum-margin object detector with a CNN to extract facial features. Another method is FaceNet, which is also based on CNN.

2. **Features Extraction:** A key step that follows the detection phase consists of extracting the features, also called the signatures, which must be sufficient to represent a human face in the form of a vector, as we can call it, encoding the face. It is necessary to examine the uniqueness and individuality of the human profile [57]. It should be mentioned that this process can be completed in the face detection step [56]. Many different techniques can do this operation, such as DNNs like ResNet [57], FaceNet, and OpenFace [72]. A novel technique was introduced in [19], an optimized hexagonal canny edge detection method. Images are first transformed to a hexagonal grid, followed by hexagonal image filtering using a gaussian filter. Then, the magnitude and direction of gradients are estimated on the three axes x, y, and z. In the next stage, non-maximum suppression is applied to the amplitude and direction of gradients. Finally, threshold selection, double thresholding, untrue edge tracking, and edge removal are done successively to reach edges in the hexagonal domain.
3. **Classification:** The encoders or the face vectors from the previous phase are classified into one of the classes offered during the training to make the general decision of either a known or unknown person. So, at this final stage, we have the process of finding an image label with a level of confidence. It is based on using supervised learning algorithms, which are frequently employed to do this type of operation [56]. At this point in the system outcome, we will be doing either identification by comparison to other faces in order to determine the person's identity or verification of the person, which entails matching one face to another for accomplishing tasks like guaranteeing access, for example.

Sometimes, detection and extraction of features can be performed simultaneously. For example, the facial features (eyes, mouth, and nose) used for feature extraction are frequently used during face detection. Hence, we have to mention the difference between handcrafted and learned features. Handcrafted features are features manually extracted by the data scientist. While, learned features are ones automatically obtained from a ML algorithm. Especially when we use DL models, like the case of CNNs that can perform the FR task completely without the intervention of any other method or the help of the engineer [56][57]. Depending on the application environment's complexity, some external factors can cause highly intra-face identity distributions (or lowly inter-face identity distributions) and degrade the accuracy of recognition. Among these factors, we can cite the database size, low or high lighting, the presence of noise or blur, disguises, partial occlusion, and certain secondary factors that are often common, unavoidable, and very challenging. In a noisy environment, image pre-processing may prove necessary [56].

Although automated FR systems must perform the three steps mentioned above [56]. Since each step is considered a critical research issue, not only because the techniques used for each step need to be improved but also because they are essential in several applications. For example, face detection is necessary to activate facial monitoring, and the extraction of facial features is crucial to identifying the person's emotional state, which is, in turn, essential in human-machine interaction [55][56][57].

3.3 Face Recognition techniques

A traditional FR system operates on photos or videos received from surveillance systems, commercial or private cameras, or similar daily hardware. In a complete automatic configuration, the system must first detect the face in the input image or video, then conduct some preprocessing steps to extract characteristics, and, lastly, identification is conducted using a good classification strategy based on the calculated characteristics [57]. Depending on the nature of the extraction and classification methods employed, FR methods are divided into four different subclasses [55][56][57], namely: (1) holistic methods, (2) local (geometrical) methods, (3) local texture descriptor-based methods, and (4) DL-based methods, as illustrated in Figure 3.3.

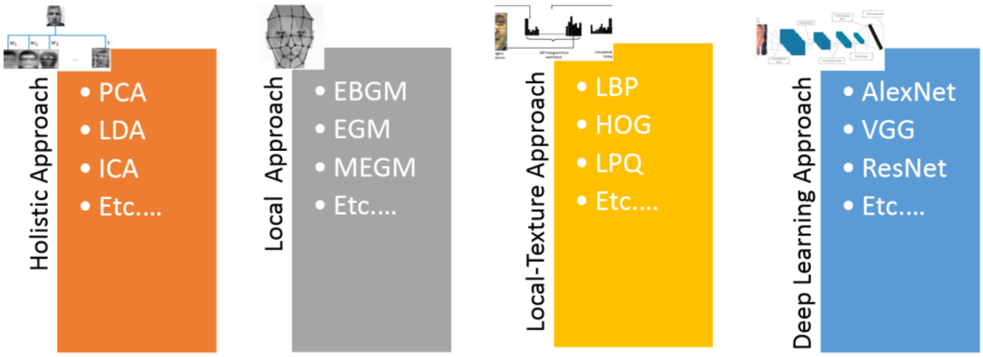


Fig. 3.3: FR Techniques [57].

3.3.1 Holistic Approach

The holistic approach or subspace-based algorithms for FR rely on using the entire face as a unique identifier rather than depending solely on specific features such as eyes, nose, or mouth. This kind of technique tries to recognize faces in an image using the whole face region as an input, treating the face as a whole. This approach harnesses the total facial structure, including the spatial relationships between distinct features and the texture of the skin, to build a holistic depiction of a person's face. It supposes that any face picture holds redundancy that can be reduced by applying the tensor's decomposition. Constructing a collection of basis vectors reflecting a lower spatial dimension (i.e., subspace) and maintaining the original set of images. In the set of basis vectors, each face in the subspace can be recreated. To assist the operation, each facial picture N×N is represented by a vector obtained by aligning the image rows. To get the non-singular basis vectors, the consequence matrix (N×N) × M is decomposed. Classification is typically done by projecting a newly obtained face picture and calculating the distance's measure with all classes described in that subspace. Besides, these methodologies may be classified into two categories, namely linear and non-linear strategies, depending on how they describe the subspace. Some famous or well-known works using this approach are: PCA, known as Eigenfaces [60], Linear Discriminative Analysis, known as Fisherface [61], and Independent Component Analysis [74] are the most common linear techniques employed for FR systems. In this technique, Eigenface is considered the pioneering and groundbreaking method. It is also known as the Karhunen-Loève expansion, primary component, or eigenvector. It is illustrated in Figure 3.4 how this technique treats a given image input.

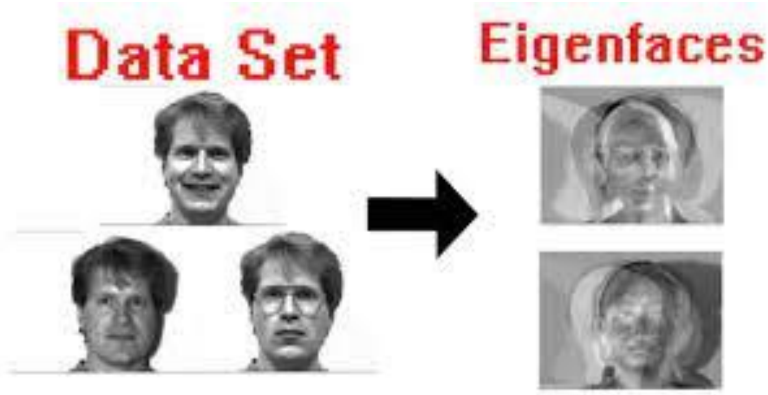


Fig. 3.4: Illustrative representation of the engine face technique [60].

3.3.2 Local Approach

The local approach to face recognition emphasizes the importance of examining specific facial traits and recognizing them as distinguishing components that contribute to an individual's unique identity. Also named geometric approach, it is based on the face landmarks. Landmarks are the primary features or information’s of a human face, used to identify and portray important portions of a face and characterize the key facial components of a given human image. The most commonly utilized landmarks on the face are the following: mouth, eyebrows, eyes, nose, jaw, the center of the iris, and ears. These landmarks offer useful information when faced with algorithms for recognition. The facial regions in the photograph do not convey the same amount of information. The forehead and cheeks, for example, have uncomplicated architecture and fewer defining patterns as compared to the nose or eyes. The landmarks in the face are used to register facial features, normalize expressions, and recognize designated positions based on the geometric distribution and the gray level pattern. The primary limitation of all approaches based on geometry is that they involve perfectly aligned facial photographs. All facial photos must be aligned to possess all reference points (e.g., mouth, nose, and eyes) displayed at the relevant place’s feature vector. The facial images are most typically manually prepared for this purpose and are often put under an anisotropic scale;

optimal automatic alignment is usually regarded as a tough task. An example of landmark extraction is presented in Figure 3.5.

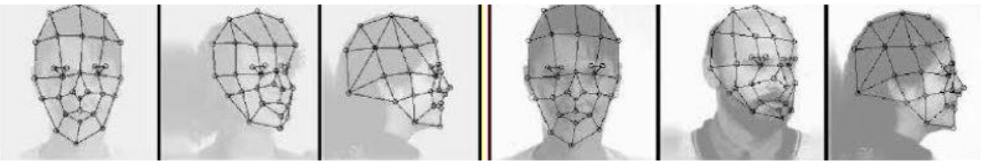


Fig. 3.5: Example of landmarks extraction using the EBGM algorithm [74].

This approach of geometry methodoloy is based on three key aspects, which are:

- **Feature extraction:** In the local approach, facial features (landmarks) are extracted individually. These features can include the eyes, eyebrows, nose, mouth, and chin. Each feature is analyzed separately to capture unique patterns and characteristics.
- **Feature based representation:** Instead of representing the entire face, local approaches build a feature-based representation. This representation is essentially a collection of distinctive features and their associated characteristics.
- **Feature matching:** During recognition, the extracted features are compared to a database of stored feature representations. A match is determined by measuring the similarity between the extracted and stored features.

3.3.3 Local Texture Approach

Feature extraction strategies focused on knowledge about the texture play a significant role in pattern recognition and CV. These based texture methods gained more attention and were introduced in many applications, such as texture classification, FR, or image indexing. They are distinctive, resilient to monotonic gray-scale changes, poor lighting, variance in brightness, and do not need segmentation. The local descriptor's goal is to transform the information at pixel-level into an appropriate form, which acquires the most compelling content insensitive to different aspects. The local descriptor's purpose is to change the information at pixel-level into an acceptable form, which acquires the most compelling content insensitive to diverse characteristics generated by fluctuations in the environment. Contrary to global descriptors that calculate features directly from the full image.

Ahonen et al. [75] (2004-2006) presented the revolutionary work of this method. they provided an innovative and successful representation of the face image based on the local texture descriptor named Local Binary Pattern (LBP). The facial picture was broken into multiple blocks, where the distributions of the LBP feature were picked and blended into an enhanced histogram used as a facial descriptor, as illustrated in Figure

3.6. The texture representation of a single area encodes the area's look, and the combination of descriptions of the entire area defines the face's global morphology.

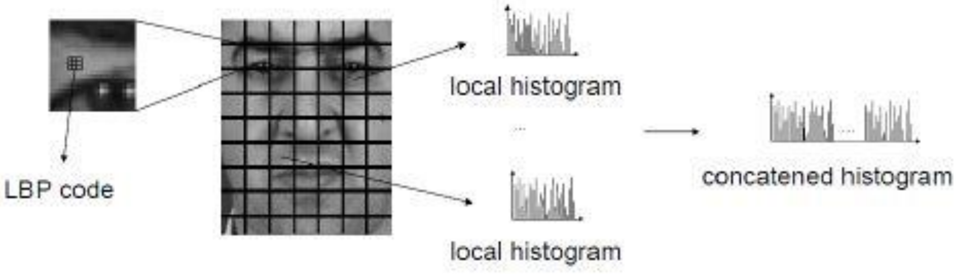


Fig. 3.6: Example of facial LBP calculation [75].

3.3.4 Deep Learning Approach

DL has won abundant contests in pattern ML and recognition during the past few years. DL, which belongs to a ML class, employs consecutive hidden layers of information-processing levels, is hierarchically organized for representation or pattern classification, and features learning. It has proven to be a game-changer in various domains, and FR is no exception. Traditional CV methods struggled with variations in lighting, pose, expression, and occlusion. DL models, on the other hand, excel at capturing complex patterns and features in data, making them ideal for FR. Various variants of neural networks have been developed in the last few years, such as CNN and recurrent neural networks (RNN), which are very effective for image detection and recognition tasks. CNNs are a very successful deep models and are been used today in many applications. Several state-of-the-art models have emerged, pushing the boundaries of FR accuracy. Here are a few notable ones:

- **VGGFace:** The VGGFace model, based on the VGG16 architecture, is a widely used baseline for FR. It employs a deep CNN to extract features from faces and achieved remarkable accuracy in early benchmarks.
- **FaceNet:** Developed by Google, FaceNet employs a triplet loss function to learn a feature representation space where faces from the same individual are close together, while faces from different individuals are far apart. This model sets new standards for FR accuracy.
- **DeepFace:** Facebook's DeepFace model employs a deep CNN with over 120 million parameters, making it one of the most powerful FR models. It uses a sophisticated pipeline to align faces and has achieved impressive results in benchmark datasets.
- **ArcFace:** ArcFace introduced the concept of additive angular margin loss to improve the discrimination power of feature embeddings. This innovation led to substantial accuracy improvements and robustness against variations in lighting and pose.

These state-of-the-art DL models have consistently outperformed their predecessors and have made significant strides in terms of accuracy and real-world applications in the FR task. These models have demonstrated accuracy rates of over 99% on benchmark datasets like LFW and MegaFace.

The remarkable thing is that all the best DL models to do FR tasks are based on CNN. CNN models are the workhorses of DL in this context. These networks are designed to automatically learn hierarchical representations of images, from low-level edges and textures to high-level facial features, making them well-suited for the intricacies of FR. Figure 3.7 illustrates the application of the CNN model for FR of a given image.

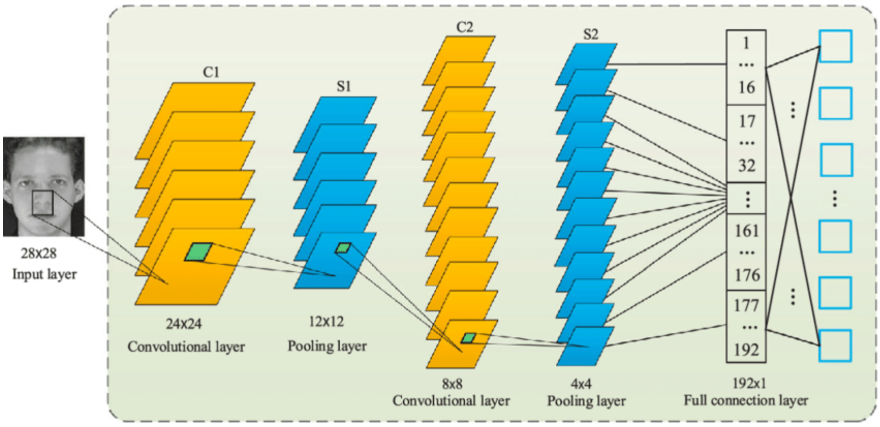


Fig. 3.7: Illustration of CNN model for FR task [75].

From a structural point of view, CNNs are made up of three main types of layers: convolution layers, pooling layers, and fully connected layers. The convolutional layer is also called the feature extractor layer because features of the image are extracted within this layer. Convolution preserves the spatial relationship between pixels by learning image features using small squares of the input image. The input image is convoluted by employing a set of learnable neurons. This produces a feature map or activation map in the output image, after which the feature maps are fed as input data to the next convolutional layer. The convolutional layer also contains rectified linear unit activation to convert all negative values to zero. This makes it very computationally efficient, as few neurons are activated each time. The pooling layer is used to reduce dimensions, with the aim of reducing processing time by retaining the most important information after the convolution opeartion. This layer basically reduces the number of parameters and computations in the network, controlling overfitting by progressively reducing the spatial size of the network. There are two operations in this layer: average pooling and maximum pooling. Average pooling takes all the elements of the sub-matrix, calculates their average, and stores the value in the output matrix. Max-pooling searches for the highest value found in the sub-matrix and saves it in the output matrix. In the fully-connected layer, the neurons have a complete connection to all the activations from the previous layers. It connects neurons in one layer to neurons in another layer. It is used to classify images into different categories by training.

3.4 Face Recognition Benchmarks

To evaluate and compare the verification or identification performance of a pattern recognition system in general and a biometric recognition system in particular, image benchmark datasets of acceptable topic size must be accessible to the public. In this part, we would like to summarize adequate datasets for testing the performance of face verification and identification systems that have been used to compare and validate FR systems, which may also be freely downloaded or certified.

- **ORL dataset:** Between 1992 and 1994, the ORL (Olivetti Research Laboratory) dataset [76] was generated at the Cambridge University Computer Laboratory. It was utilized in the framework of a FR project carried out in partnership with the Speech, Vision, and Robotics Group of the Cambridge University Engineering Department. It contains 400 frontal facial photos for 40 persons with different facial emotions (open/closed eyes, smiling/not smiling), conditions of illumination, haircuts with or without the beard, mustaches, and spectacles. All the samples were taken against a black uniform background with the subjects in a vertical frontal stance (with a little side mobility). Each sample is 92×112 pixels with 256 grayscale pictures, with tilting tolerance and up to 20° rotation, as illustrated in Figure 3.8.

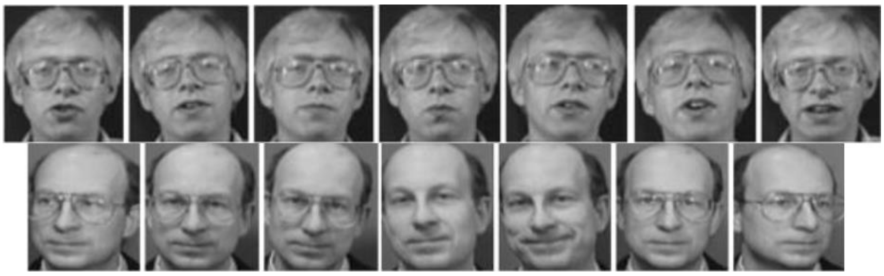


Fig. 3.8: ORL database samples[76].

- **FERET dataset :** The FERET dataset [77] was created in the Department of Defense Counterdrug Technology Development Program Office between 1993 and 1996. The goal was to develop FR capabilities through machines that could be used to assist in authentication, security, and forensic applications. The facial images were acquired in 15 sessions in a semi-controlled environment. It covers 1564 sets of facial images for 14,126 images that comprise 1199 people and 365 duplicate sets of facial images. A duplicate set is another set of facial images of an individual that already exists in the database and was typically collected on another day. An illustration is presented in Figure 3.9.

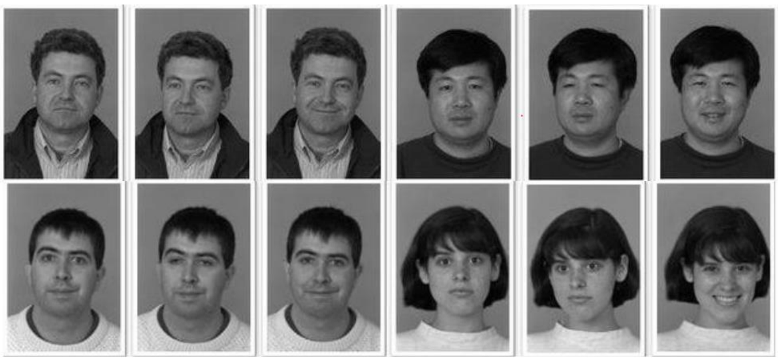


Fig. 3.9: FERET database samples [77].

- **AR dataset :** Martinez and R. Benavente established the database AR (Alex and Robert) [78] in 1998 at the CV Center, Barcelona (Spain). The database includes more than 3000 colored facial images of 116 participants (53 women and 63 men). This database's photos were taken under different illumination settings, including frontal views, facial expressions, and occlusions (by scarf and sunglasses). All photographs were acquired in strictly controlled situations; no wear requirements (glasses, clothes), hairstyle, or makeup were required for contributors. Each individual took part in two distinct sessions, spread over two weeks. So, there are 26 photographs from each theme. Both sessions acquired the same photos. The generated RGB facial images are 768×576 pixels in size. The database provides 13 kinds of images as represented in Figure 3.10, which are: (1) neutral expression; (2) smile; (3) anger; (4) screaming; (5) left light on; (6) right light on; (7) both side-lights on; (8) wearing sunglasses; (9) wearing sunglasses and left light on; (10) wearing sunglasses and right light on; (11) wearing a scarf; (12) wearing a scarf and left light on; and (13) wearing a scarf and right light on.

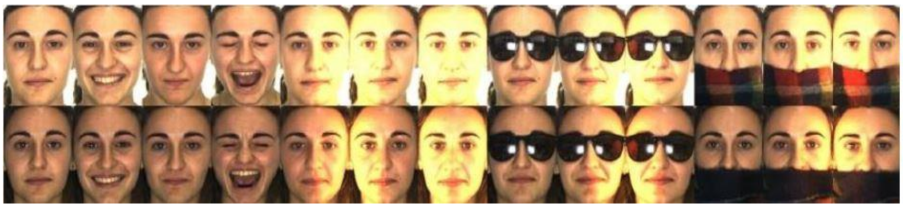


Fig. 3.10: AR database samples [78].

- **LFW dataset :** The facial image database LFW created by Huang et al. [79] in 2007, desired to study the unconstrained problem of FR, such as variation in posture, facial expression, race, background, ethnicity, lighting, gender, age, color saturation, clothing, camera quality, hairstyles, focus, and other parameters (Figure 3.11). It contains 13,233 facial photographs taken from the web by 5749 individuals, of which 1680 have two or more distinct images. Each image is 250×250 pixels in size. Two protocols are defined for using the training data: image-restricted and unrestricted training. Under Image-Restricted, the identities of images are ignored in training. The key distinction is that the unrestricted protocol lets you construct as many impostors and real pairs as feasible over the restricted training pairs. Three types of aligned images are proposed: (1) the funneled images; (2) LFW-A employed an unreleased approach for alignment; and (3) deep funneled images [28].

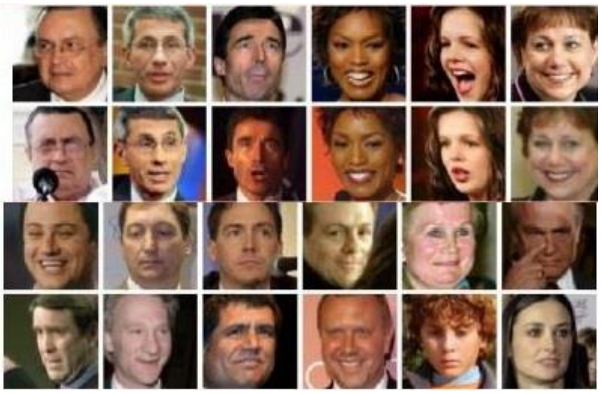


Fig. 3.11: LFW database samples [79].

- **CASIA-WebFace dataset :** Another large-scale dataset for the FR task, named CASIA-WebFace, was selected from the IMDb website, containing 10,575 individuals and 494,414 facial photos. It was constructed in 2014 by Yi et al. [80] at the Institute of Automation, Chinese Academy of Sciences (CASIA). Some of its samples are shown in Figure 3.12.

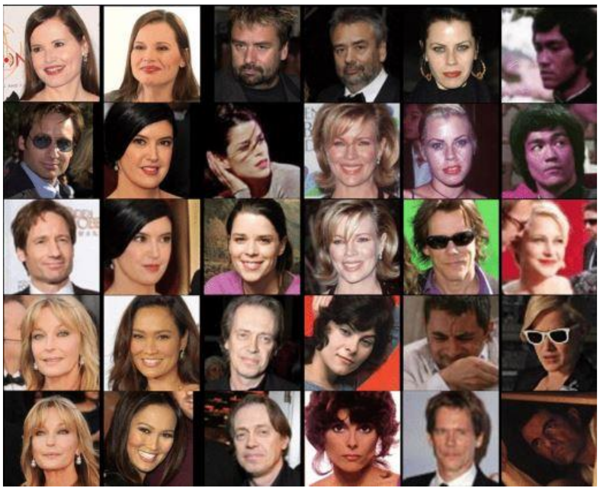


Fig. 3.12: CASIA-WebFace database samples [80].

- **VGGFACE database :** VGGFACE (Visual Geometry Group) is a large-scale training database that is assembled from the internet by merging automation and people in the loop. It comprises 2.6M photos, over 2.6K identities (Figure 3.13). It was generated at the University of Oxford in 2016 by Parkhi et al [67].

Fig. 3.13: VGGFACE database samples[67].

- **LFR dataset :** LFR (Left-Front-Right) is a face recognition dataset presented by Elharrouss et al. [81] from Qatar University in 2020 to overcome pose-invariant FR in the wild. Pose variation reveals a tough FR problem in an unrestricted context. To deal with this issue, a CNN model for estimating pose is proposed. This model is trained using a self-collected dataset generated from three standard datasets: LFW, CFP, and CASIA-WebFace, employing three classes of face image capture: left, front, and right. A dataset of 542 IDs is therefore constructed, representing each subject's photos on the left, front, and right faces. Each folder on the left and right has 10–100 facial photographs, while the front folder contains 50–260 images. An illustration is presented in Figure 3.14.

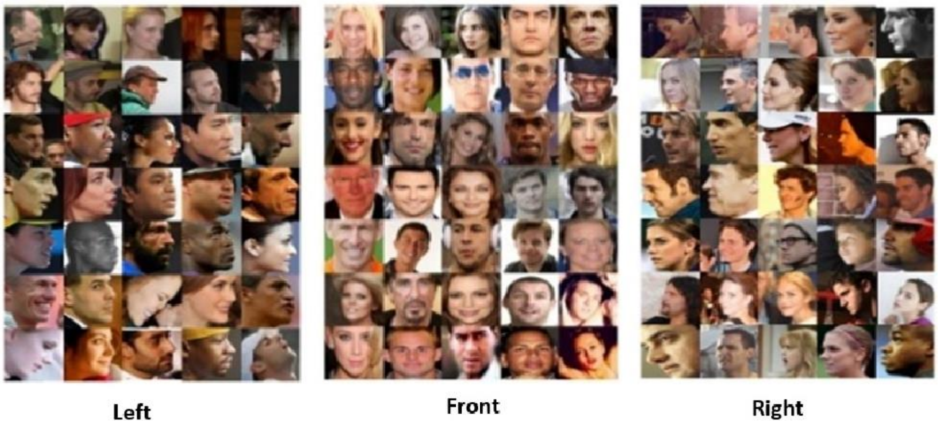


Fig. 3.14: LFR database samples [81].

- **RMFRD and SMFRD:** Masqued FR dataset During COVID-19, practically everyone wears a mask to restrict its spread, making standard FR technologies inefficient. Hence, enhancing the recognition performance of the current FR technology on masked faces is quite significant. For this reason, Wang et al. [82] presented three types of masked face datasets, which are:

1. **Masked Face Detection Dataset (MFDD)[MFDD]:** it may be utilized to train a masked face detection model with precision.
2. **Real-world Masked Face Recognition Dataset (RMFRD):** it contains 5,000 photographs of 525 persons wearing masks and 90,000 pictures of the same 525 individuals without masks gathered from the Internet (Figure 3.15.a).
3. **Simulated Masked Face Recognition Dataset (SMFRD):** in the interim, the proposers applied different techniques to place masks on the typical large-scale facial datasets, such as the LFW and CASIA WebFace datasets, thereby enhancing the volume and variety of the masked FR dataset. The SMFRD dataset comprises 500,000 facial photos of 10,000 humans, and it can be deployed in practice with their original unmasked counterparts (Figure 3.15.b).

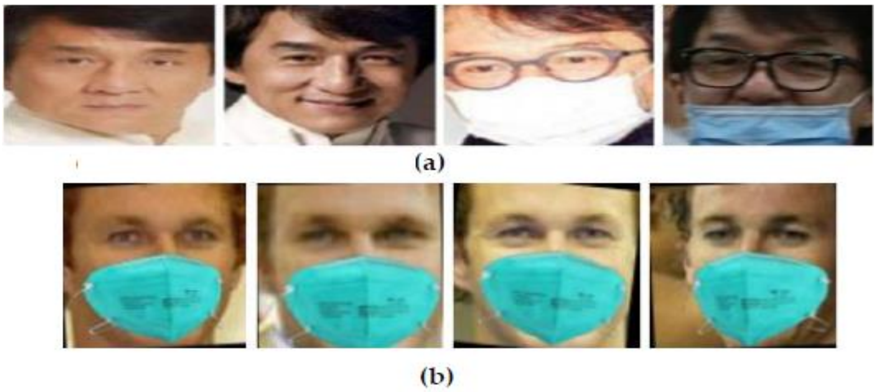


Fig. 3.15: RMFRD and SMFRD databases samples[82].

3.5 Face Recognition for Smart City Surveillance a Literature Review

This section delineates diverse scholarly endeavors put forth by researchers as potential resolutions for FR systems through the utilization of ET and IoT technologies. DL-based intelligent FR in an IoT-cloud environment is proposed in [83]. This research suggests a tree-based DL for automatic FR in a cloud setting. As mentioned in the paper, the proposed deep model is computationally less expensive without affecting accuracy. In the model, an input volume is split into numerous volumes, and a tree is produced for each volume. A tree is described by its branching factor and height. Each branch is represented by a residual function, which is made of a convolutional layer, a batch normalization, and a non-linear function. The proposed model is examined in standard databases. A comparison of performance is also done with state-of-the-art deep models for FR. The results of the experiments reveal accuracies of 98.65%, 99.19%, and 95.84% on the FEI, ORL, and LFW databases, respectively. The authors in [84] proposed a smart Ssecurity system using FR on Raspberry Pi. In this study, the HC algorithm was used for face detection, combined with the Eigenfaces approach for face identification. The algorithm was trained using more than 50 different face photographs of 10 different individuals recognized as known users. A second training phase used more than 400 data images from the ORL database that were recognized as unknown users. This security solution was built with an Arduino Uno board included to ensure the integration of some sensors that the RPi does not support and to generate the pulse width modulation signals necessary to control other hardware modules inside of this project with a connection to an RPi camera device. The RPi microcontroller served as the system's brain. The system displays a 95.5% accuracy rate. In [85] the researchers proposed a smart home security using facial authentication and mobile applications. In this work, a system based on the RPi module with a connected RPi camera was employed for home automation security to control door entry. using the LBPH algorithm to recognize faces and the Haar-Cascade technique for detection. After training on more than 100 photos for 7 people, the system had an accuracy of 92.86%. The article [86] presents a novel IoT-based face mask detection system for public transportation, especially buses. This system would collect real-time data via FR. The main objective of the paper is to detect the presence of face masks in a real-time video stream by utilizing DL, ML, and image processing techniques. To achieve this objective, a hybrid DL and ML model was designed and implemented. The model was evaluated using a new dataset in addition to public datasets. According to the authors, the results showed that the CNN classifier has better performance than the Deep Neural Network classifier; it has almost complete face-identification capabilities with respect to people’s presence in the case where they are wearing masks, with an error rate of only 1.1%. Overall, compared with the standard models, AlexNet, Mobinet, and the You Only Look Once (YOLO) method, the proposed model showed better performance. Moreover, the experiments showed that the proposed model can detect faces and masks accurately with low inference time and memory, thus meeting the IoT's limited resources. The proposed system supports real-time processing of the inputs and forces people to wear the face mask as per the guidelines. The proposed CNN model helped us achieve a high accuracy of 99%. Other work presented by the authors in [87] suggesting an edge FR System based on one-shot augmented learning where they combined a number of different algorithms. The combination that produced the best results was using HOG for face detection, ResNet for face encoding, and K-NN for classification. In this research, they obtained face data images with an ESP32-CAM attached to a passive infrared sensor, then sent the image to an RPi for processing purposes and to determine whether the person was recognized or not. A notification will be sent to the Telegram account of the system administrator. This project shows an accuracy result of 94.77%. Another solution proposed in [88] is an application for automating the household doorbell, which can not only solve security issues but also offer extra flexibility to smart house control by recognizing the person at the doorstep and announcing their name. They propose a system by making use of a Raspberry Pi with an ARMv8 Cortex-A53 core. The project aims at porting the OpenCV library to the Raspberry Pi board and using its pre-trained classifier Haar-Cascade and recognizer Local Binary Pattern Histogram (LBPH) for face detection and recognition purposes.

Furthermore, using the Raspberry Pi accelerated command-line media player, the name of the person will be announced, and faces unknown to the database will be captured and stored. In [89], the authors of the paper propose a sparse individual low-rank component-based image representation. Such representations of testing images can be based on individual subjects' low-rank components. Theoretically, they put the L2-normmalization constraint on intra-subject coefficients to represent testing images, thus making intra-subject coefficients dense. Hence, they alleviate the impact of an under-sampled training dataset and its same inter-subject variation on classification performance. According to them, they solve a convex minimization problem in polynomial time via an augmented Lagrange multiplier scheme to get the solution of the proposed SILR. The scheme can reduce the influences from the same inter-subject variation and contribute to an accurate recognition of the under-sampled training dataset. We adopt sparse individual low-rank component representation and minimum reconstruction residual to recognize testing images. Results from face-to-face benchmark databases and many non-standard datasets show that the proposed SILR is better than the other state-of-the-art methods for FR, with recognition rates of 82.98%, 61.11%, 89.15%, and 94.91% on the LFW, CelebA, AR, and CMU-PIE databases, respectively. The authors in [90] suggested a home security system with FR based on CNN. In this study, a CNN was suggested for the recognition task. It was trained using the HC technique for face detection on 1048 facial data photos of homeowners. They were able to attain 97.50% accuracy with their project, a door locking system based on RPi 3, connected to an RPi camera module. In [91] a solution was proposed of an IoT with DL-based FR approach for Authentication in control medical systems". In this work, three distinct approaches were suggested, all of which utilized the Haar cascade for face detection. Following a training process using 8422 collected face images of 100 different people, they used the pretrained ResNet-50 model and SVM algorithm in the first approach, a pretrained VGG model and SVM algorithm in the second approach, and the Linear Binary Pattern Histogram algorithm in the third approach, achieving the following accuracy results, respectively, of 99.56%, 98.49%, and 98.47%. An RPi camera module and RPi 4 board are included in the project's smart door locking control via face authentication in a medical system. The authors in [92] suggested an IoT-based FR system for home security using LBPH algorithm. This project was a web camera module-based RPi 3 solution that used the HOG approach to detect faces and the LBPH methodology for recognition. A local dataset of 100 photos was used for training, and 40 images were used for testing. An accuracy of 90.00% was recorded. An other work [93] proposing a smart intelligence mirror system. The RPi and an RPi camera were utilized for this system. The LBPH used for FR was trained on over 100 photos from a local database. The method uses a decision tree algorithm for face detection and the Gini index as a facial classifier. The obtained accuracy was around 87.00%. The researcher in [94] suggested a smart home security system using IoT, FR, and Raspberry Pi. In this study, they suggested a method for an automatic door locking control system using LBPH for recognition. Based on the RPi camera that has been connected to the RPi microcontroller for face recording, it will be recognized using HOG. This mythology had a 90.00% accuracy rate. A framework using visual recognition for IoT-based SC surveillance was proposed in [95]. In this manuscript, a quick subspace decomposition over Chi Square transformation is proposed. This technique extracts the characteristics of visual data using a LBPH. The redundant features are removed by using rapid subspace decomposition over the Gaussian-distributed LBP features. As mentioned in the paper, the redundancy removal is a major contribution to memory and time consumption for battery-powered surveillance systems, which makes the methodology suitable for all image recognition applications and deployment into IoT-based surveillance devices due to improved dimension reduction. The technique was implemented on the Raspberry Pi and validated over well-known databases. The least error rate is attained by the suggested technique with maximal feature reduction in minimum time, with 0.28%, 1.30%, 9.00%, and 2.5% rates over AR, O2FN, LFW, and Dyn Tex++ databases, respectively. In [96], the authors aim to build a low-cost intelligent mirror system that can display a variety of details based on user recommendations. Therefore, in this article, IoT and AI-based smart mirrors are introduced that will support the users in receiving the necessary daily updates of weather information, date, time, calendar, to-do list, updated news headlines, traffic updates, COVID-19 case status, and so on. Moreover, a face detection method was also implemented with the smart mirror to make the architecture more secure. Their proposed application, named MirrorME, provides a success rate of nearly 87% in interacting with the features of FR and voice input. The proposed workflow consists of detecting and recognizing faces using HOG and linear SVM. The mirror is capable of delivering multimedia facilities while maintaining high levels of security within the device. The work [97] addresses the application of DL models for anomaly detection and FR in IoT devices within the context of smart homes. Six models, namely, LR-XGB-CNN, LR-GBC-CNN, LR-CBC-CNN, LR-HGBC-CNN, LR-ABC-CNN, and LR-LGBM-CNN, were proposed and assessed for their performance. The models were trained and evaluated on labeled datasets of sensor readings and face photos, employing a range of performance criteria to assess their usefulness. Performance evaluations were done for each of the proposed models, indicating their strengths and opportunities for development. According to their comparison analysis of the models, the LR-HGBC-CNN model consistently beat the others in both anomaly detection and FR tasks, obtaining high accuracy, precision, recall, F1 score, and AUC-ROC values. For anomaly detection, the LR-HGBC-CNN model achieved an accuracy of 94%, a precision of 91%, a recall of 96%, an F1 score of 93%, and an AUC-ROC of 0.96. In FR, the LR-HGBC-CNN model achieved an accuracy of 88%, precision of 86%, recall of 90%, F1 score of 88%, and an AUC-ROC of 0.92. According to the authors, the models displayed promising capabilities in identifying anomalies, recognizing faces, and integrating these functionalities into smart home IoT devices. Table 3.1 summarizes the mentioned related works, each with its own key methods used for face detection, feature extraction, and face recognition, the utilized IoT gadgets, and the obtained accuracy.

Table 3.1: Summary of previous related works.

Work	Methods			IoT Based Modules	Accuracy
	Face Detection	Features Extraction	Face Recognition		
[90, 2020]	Haar-cascade	CNN	CNN	RPi + RPi Camera	97.50 % Local database
[84, 2019]	Viola-Jones	PCA	Eigenface	RPi + Arduino-Uno + RPi Camera	95.00 % ORL database
[87, 2022]	HOG	ResNet	K-NN	ESP32-CAM + RPi	94.77 % Local Database
[91, 2018] 1st Appraoch	HC	ResNet-50	SVM	RPi + RPi Camera	99.56 % Local database
[91, 2018] 2nd Appraoch	HC	VGG-16	SVM	RPi + RPi Camera	98.49 % Local database
[91, 2018] 3rd Appraoch	HC	LBPH	LBPH	RPi + RPi Camera	98.47 % Local database
[92, 2018]	HOG	PCA	LBPH	RPi + RPi Camera	90.00 % Local database
[85, 2022]	HC	LBPH	LBPH	RPi + RPi Camera	92.86 % Local database
[93, 2021]	Decision-tree	PCA	LBPH	RPi + RPi Camera	≈ 87.00 % Local Database
[90, 2020]	HOG	LBPH	LBPH	RPi + RPi Camera	90.00 % Local database
[83, 2020]	Tree-based deep model			IoT Devices + Cloud Storage	98.65%, 99.19%, and 95.84% FEI, ORL, and LFW databases, respectively
[95, 2021]	Subspace Decomposition Over Chi Square	LBPH	LBPH	RPi + RPi Camera	≈ 99.00 %, 97.00%, 93.00%, 98.00% AR, Dyn Tex, LFW and OF2N databases, respectively

[86, 2022]	Single Shot Multibox Detector	CNN	CNN	RPi + RPi Camera	99.00 % Local dataabse
[88, 2021]	HC	LBPH	LBPH	RPi + RPi Camera	Not Mentioned
[89, 2021]	Sparse Individual Low-rank component-based image representation			Iot Application	82.98%, 61.11%, 89.15% and 94.91% on LFW, CelebA, AR and CMU PIE
[96, 2021]	HOG and linear SVM			RPi + RPi Camera	87% in interacting with the features of FR and voice input
[97, 2023]	LR-HGBC-CNN			IoT gadgets	94% on Real and Fake Face Detection database

3.6 Conclusion

This chapter provides a thorough assessment of the FR domain and its intricate connection to SC surveillance. Starting with a basic introduction that elucidates the essence of FR and unravels the intricate procedures involved in its deployment as a surveillance system committed to improving SC security. Diving deeper into the technical aspects, the chapter exhaustively explains a range of FR techniques, including holistic, local, local texture, and DL approaches. This comprehensive analysis gives readers a bird's-eye perspective of the various approaches that contribute to the usefulness of FR systems in surveillance, demonstrating the breadth of instruments available for improving security measures. The narrative easily shifts from the issue of benchmarking to the aim of building solid FR systems. The chapter walks through the key benchmarks used to confirm the effectiveness of built FR systems, providing a critical lens through which the reliability and performance of these systems can be evaluated. A rigorous literature analysis is offered as a poignant end to the chapter, showing the numerous contributions of FR in ISS as feasible options for strengthening security within the SC surveillance area. This retrospective analysis not only consolidates the key insights gleaned from the preceding sections, but also situates the current discourse within the larger landscape of existing research, emphasizing the ongoing evolution and impact of FR technologies in addressing contemporary society's security challenges.

Chapter 04 :
Our Contribution

The importance of surveillance systems is growing since they are related to person safeguarding, public safety, and effective urban management. As smart cities continue to develop, surveillance systems provide new opportunities and benefits through the use of ET, IoT, and advanced ML approaches and methods. Providing solutions that merge IoT and ET into the surveillance aspect is the need of the hour and what the researchers and scientists are striving to do, characterized by high efficiency and an inexpensive cost. This part of the chapter will highlight the different scientific contributions that arose as a result of the study for this thesis.

4.1 Smart City Surveillance: Edge Technology Face Recognition Robot Deep Learning Based

Starting with the proposed solution to our main studied problem, which was a smart surveillance robot using FR based on ET and DL models, This scientific contribution [103] was a suggested solution for SC surveillance aspect, and it was built using several approaches and methods that will be discussed in this part.

4.1.1 Methodology

The proposed system in this study was created following various steps, as depicted in Figure 4.1. The methodology was divided into two main points: the hardware building the edge car robot, and the software, which was the FR system.

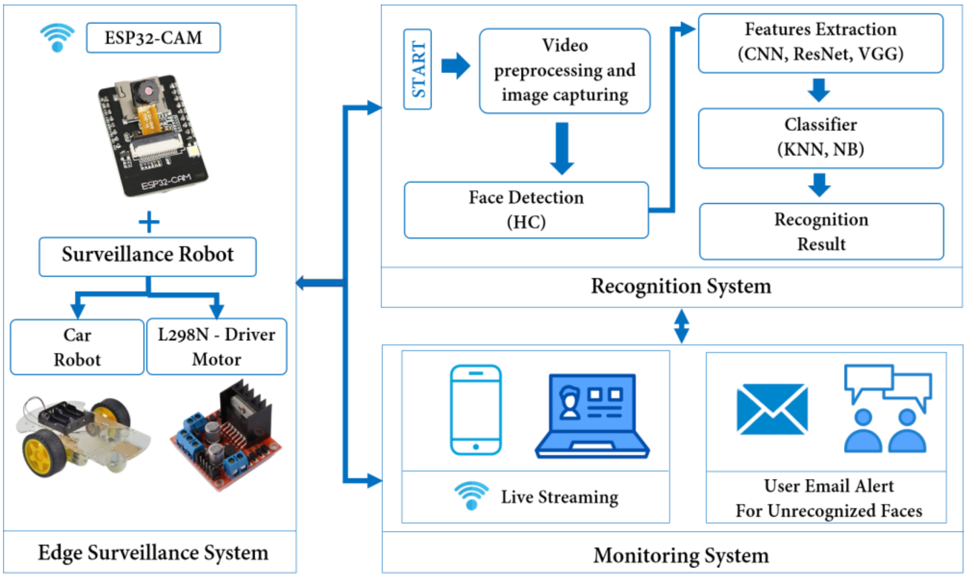


Fig. 4.1: The proposed edge surveillance security system [103].

The hardware side of the suggested solution is predicated on the two aspects. The first one is the edge surveillance system and the second is the monitoring system. Its functionality is as follows:

- Step 1: The edge surveillance robot captures the area with the camera module.

- Step 02: The identification system takes the video supplied by the robot, preprocesses it, and extracts the images in real time.
- Step 03: If a face is detected, the model will take its place and apply all the necessary processes to recognize the face.
- Step 4: Presenting the findings on the live streaming application (phone or desktop) and also notification email alerts will be provided in case of unrecognized faces.

The edge robot was created employing many components:

- ESP32-CAM module.
- L298N Motor Driver
- Car robot that comes with 1 car chassis, 2 DC gear motors, 2 car tires, 1 universalwheel, 1 battery holder, and jumper wires.

Starting with the main module, the ESP32-CAM, which is a tiny microcontroller and a low-cost ESP32-based development board featuring an embedded camera, That works autonomously, powered by its own Wi-Fi and Bluetooth connectivity, among other specifications, as follows:

- Built in 520 KB of SRAM plus 4 MB of PSRAM.
- with SD card slot support.
- Built-in flash LED.
- The camera supports, for instance, an OV5640 camera with a 5-megapixel picture sensor.

The L298N module is a twin H-Bridge motor driver that can simultaneously control the speed and direction of two DC motors. We can see that both gadgets have several pins and outputs, as represented in Figure 4.2, that allow assembling the hardware of the system together based on a circuit schematic diagram as illustrated in Figure 4.3, where each pin from one module will be connected to a specific one in the other as described in Table 4.1.

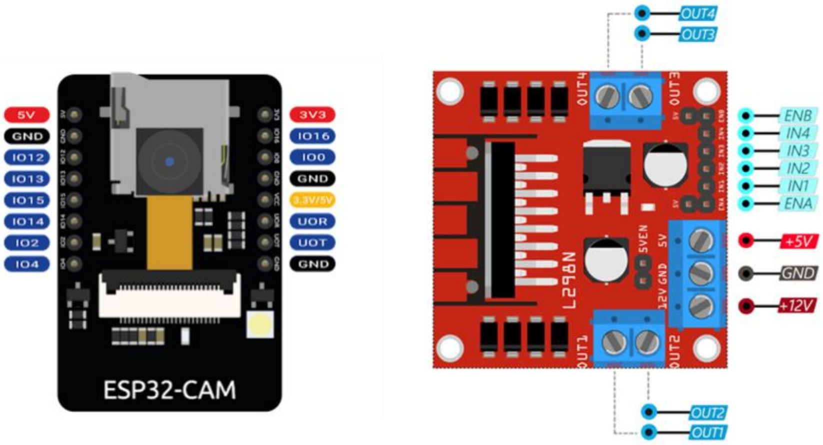


Fig. 4.2: IoT gadgests pins [103]

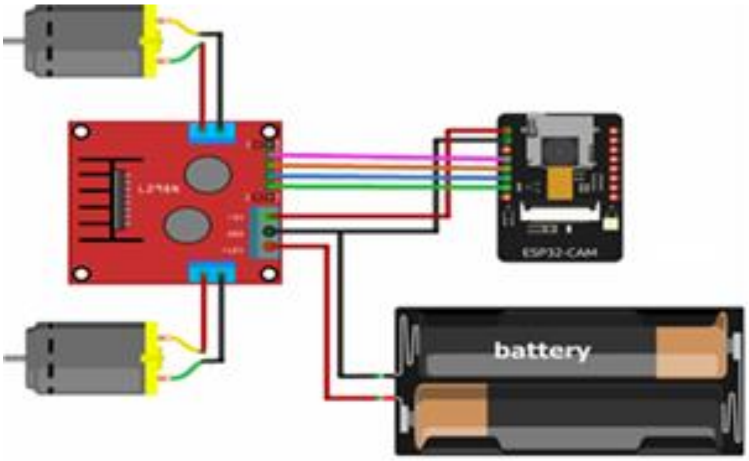


Fig. 4.3: Edge robot circuit schematic diagram [103]

Table 4.1. System pins connections.

ESP32 CAM Pins	L298N Motor Driver Pins
IO12	ENA
IO13	IN1
IO14	IN3
IO15	IN2
IO12	ENB
IO2	IN4

Other connections have been created between the two DC motors that are accountable for the wheel’s movements and the OUT-bleu terminals in the motor driver, which have three more connections, one of which is a 12V connection to the battery. The GND terminal is connected to the battery's negative terminal as well as the ESP32-CAM GND pin, while the 5V terminal is connected to the ESP32-CAM module's 5V terminal. The final result of the assembled edge surveillance car robot is illustrated in Figure 4.4.

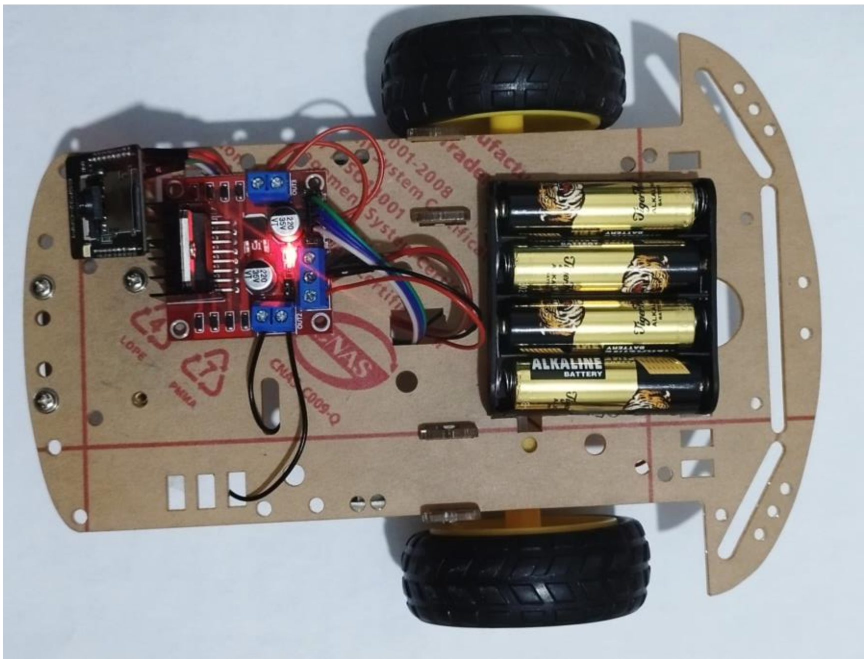
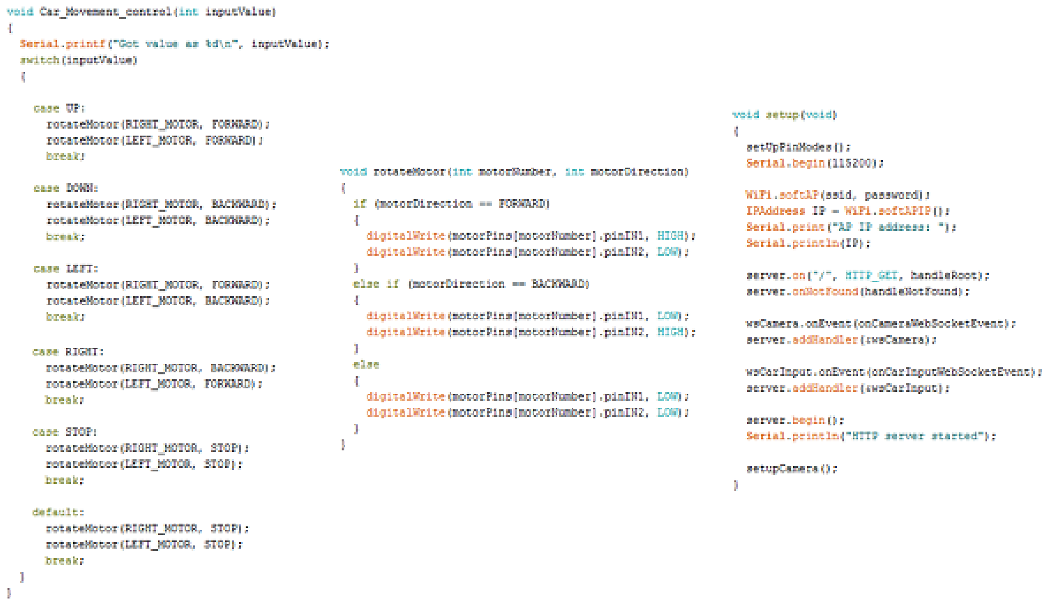


Fig. 4.4: Edge surveillance robot [103].

After putting all the pieces together, in order to make the system operational, both for live streaming via the ESP32-CAM and for controlling the auto robot via the L298N motor drive, we have to program the ESP32-CAM board in a specific way to synchronize the operations. Using the Arduino Integrated Development Environment, we upload the main code developed in C++ to complete the work. Figure 4.5 depicts the primary functions for regulating the car's movement with synchronization with the L298N motor driver. We also configured the ESP32-CAM module to be a local network provider that can be accessed using either a PC or a smart phone over the WIFI on 192.168.4.1 local IP address. Controlling and using the edge system will be done via a web application using a web browser.

Since the video streaming will be through the ESP32-CAM module, we do a preprocessing step on the acquired video to extract the facial photos, applying the same techniques as we did with the data earlier. Where each frame of the video will be transformed into an image, to detect the face inside it, then features extraction will be applied using DL models to get to the classification of the encoded features using the classification algorithms, based on what the models are trained before with a confidence threshold. As a last component of the suggested edge surveillance system, we have added the opportunity to both operate the auto robot via web application and examine the streaming live video, as represented in Figure 4.6. The automobile movements can be controlled using the buttons to go in whatever directions we desire, with speed and light regulation also. The email notification is implemented in our system to alert the user to the unknowing faces as a security response via the email address saved inside the system that will be activated if the system does not recognize people inside the surroundings that the vehicle robot is accountable for.



```
void Car_Movement_control(int inputValue)
{
  Serial.printf("Got value as %d\n", inputValue);
  switch(inputValue)
  {

    case UP:
      rotateMotor(RIGHT_MOTOR, FORWARD);
      rotateMotor(LEFT_MOTOR, FORWARD);
      break;

    case DOWN:
      rotateMotor(RIGHT_MOTOR, BACKWARD);
      rotateMotor(LEFT_MOTOR, BACKWARD);
      break;

    case LEFT:
      rotateMotor(RIGHT_MOTOR, FORWARD);
      rotateMotor(LEFT_MOTOR, BACKWARD);
      break;

    case RIGHT:
      rotateMotor(RIGHT_MOTOR, BACKWARD);
      rotateMotor(LEFT_MOTOR, FORWARD);
      break;

    case STOP:
      rotateMotor(RIGHT_MOTOR, STOP);
      rotateMotor(LEFT_MOTOR, STOP);
      break;

    default:
      rotateMotor(RIGHT_MOTOR, STOP);
      rotateMotor(LEFT_MOTOR, STOP);
      break;
  }
}
```

```
void rotateMotor(int motorNumber, int motorDirection)
{
  if (motorDirection == FORWARD)
  {
    digitalWrite(motorPins[motorNumber].pinIN1, HIGH);
    digitalWrite(motorPins[motorNumber].pinIN2, LOW);
  }
  else if (motorDirection == BACKWARD)
  {
    digitalWrite(motorPins[motorNumber].pinIN1, LOW);
    digitalWrite(motorPins[motorNumber].pinIN2, HIGH);
  }
  else
  {
    digitalWrite(motorPins[motorNumber].pinIN1, LOW);
    digitalWrite(motorPins[motorNumber].pinIN2, LOW);
  }
}
```

```
void setup(void)
{
  setUpPinModes();
  Serial.begin(115200);

  WiFi.softAP(ssid, password);
  IPAddress IP = WiFi.softAPIP();
  Serial.print("AP IP address: ");
  Serial.println(IP);

  server.on("/", HTTP_GET, handleRoot);
  server.onNotFound(handleNotFound);

  wsCamera.onEvent(onCameraWebSocketEvent);
  server.addHandler(&wsCamera);

  wsCarInput.onEvent(onCarInputWebSocketEvent);
  server.addHandler(&wsCarInput);

  server.begin();
  Serial.println("HTTP server started");

  setupCamera();
}
```

Fig. 4.5: ESP32 CAM borad main functions [103].

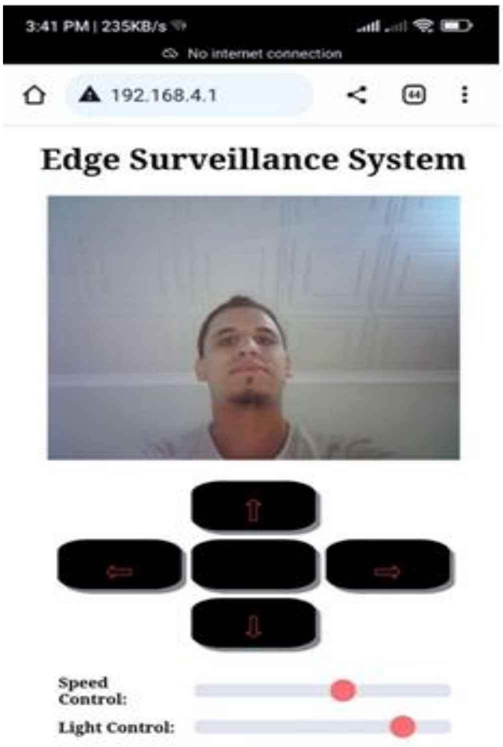


Fig. 4.6: The proposed edge surveillance system web application [103].

Moving to the presentation of the main points of building the FR system, which indicate the software side of our solution, which is based on data preparation and pre-processing, face detection, feature extraction, and classification to reach the finale model, which will be a combination of multiple methods and DL models.

The data section is divided into two parts: one for the benchmarks to validate our FR model, choosing the LFW, the ORL, and the AR databases. While the other is for a local dataset to test the surveillance system itself. In order to build this local database and to obtain a successful, correct conclusion, a considerable amount of training data is required, but on the other hand, it is a difficult and challenging job to complete in a real-time setting. Therefore, we develop our own database utilizing the data augmentation approach. It is applied to generate extra and fresh samples by changing the existing data images. Therefore, we blended three separate human face photographs together; two of them were chosen randomly from the AR database, and the third was from the author themselves. The augmentation procedure is made by applying different operations as defined and indicated in Table 4.2.

Table 4.2. Data augmentation operations.

Parameter	Interval or Value
Zoom	[-5, 5]
Shear	0.5
Height shift	0.3
Width shift	0.3
Rotation	[0, 15]

Using this method, we generated a local database with three lables including 2430 face photos with 810 data images per class.

Reshaping all the photographs to 64 by 64 grayscale photos is another preprocessing step to assure that all the data have the same shape. Next, a normalization approach will be applied to check that all the data are in the same range.

After preparing four separate datasets, a second step of face detection was done utilizing the HC algorithm. This algorithm uses a variety of predetermined features built on the foundation of Haar-like features, including edges, lines, and corners, to identify face features like the eyes, nose, and mouth. It accepts the frontal face as an input that will be subjected to a cascade application of the classifier. The image is separated into pieces before the classifier is applied to each area. If a region is recognized as a face, the cascade continues to the next phase; otherwise, it is discarded, and the next region is processed.

Our proposed FR model is built on CNNs since it has the power to emulate the human visual system's hierarchical processing. CNNs train and extract internal representations or features straight from raw picture data, making them particularly successful for tasks like FR, image categorization, object identification, and more. In this portion, we will cover the training procedures for our DL models. Three alternative DL models will be applied; two of them are predefined: the VGG and the ResNet model.

The third one is a proposed CNN developed using the TensorFlow [30] and Keras [31] frameworks. Our suggested CNN is made of 13 layers and has an architecture as indicated in Table 4.3.

Table 4.3. Proposed CNN model architecture.

Layer Type	Number of Layer
Two-Dimensional Convolutional Layer (Conv2D)	04 layers
Batch Normalization Layer	02 layers
Max Pooling Layer	02 layers
Dense Layer	03 layers

Following normalization, feature extraction and redundancy reduction take place in a pair of convolutional layers paired with pooling layers. Simple features come together efficiently over time. The features are then partially blended, with each resulting feature accounting for a fraction of the configuration of the selected class. Finally, the fully connected layer receives these top qualities and offers an estimation of the categorization.

4.1.2 Expirements and Results

The experiments made in this work, from training, testing, and validating the models, were using Google Colab [33], the free tier of the platform that delivers the demands with a system setup as given in Table 4.4.

Table 4.4. Google Colab platform free tier configuration.

Hardwar	Specifications
GPU	Nvidia Tesla K80
CPU	2x Intel Xeon @ 2.20GHz
RAM	13 GB
HARD DISK	78 GB

The model's performance evaluation was based on the accuracy metric as described in (1).

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

Where: TN (true negative) are the examples correctly classified as negative class; TP (true positive) are the examples correctly classified as positive class; FP (false positive) are the examples incorrectly classified as positive class; and FN (false negative) are the examples incorrectly classified as negative class.

As the first step in presenting the results of our study, we will start by showing the model's validation over the benchmarks. As seen in Figure 4.7, validating the models over the LFW, AR, and ORL databases yields various findings in terms of accuracy.

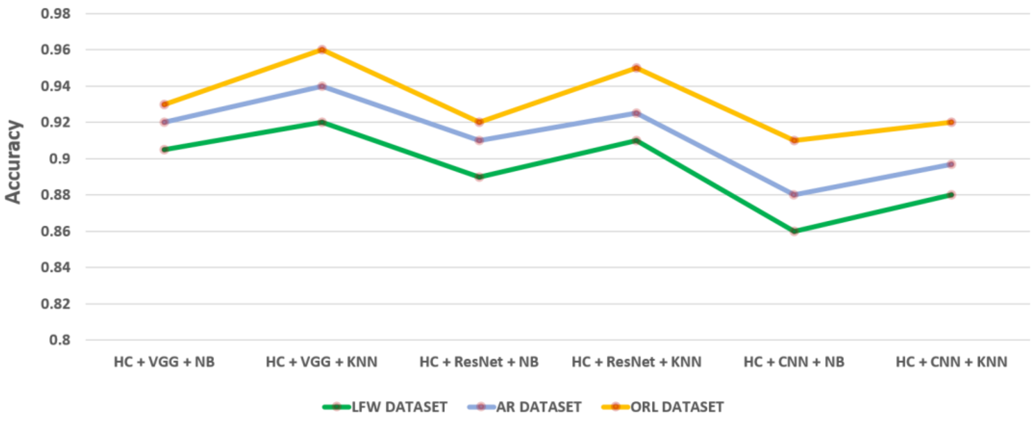


Fig. 4.7: Models validation over the benchmarks [103].

For the LFW Database, the training set was 90% from this database, while the remaining was for the testing set with 1330 samples using the train-test cross-validation. Each one of the model’s combinations has its own accuracy. We can observe from Figure 12 that the best model with the highest accuracy over this database is HC + VGG + KNN, with a value of 92.00%. While for the AR Database, we conduct cross-validation to get 10% of the data for testing using 400 samples. Analyzing the results of each combination displayed in Figure 12, it is seen that the HC + VGG + KNN model combination fit the best, with a 94.00% accuracy rate. Finally, for the ORL database The results illustrate the accuracy of each combination in this database using a testing set of 40 samples with the same percentage of 10% from the complete database. Also, we notice that the HC + VGG + KNN model combination is the best, with 96.00% accuracy. As a validation results summary, we conclude that the HC + VGG + KNN combination delivers the maximum best accuracy rate.

In order to describe the proposed surveillance system results, we first show the best model that was picked for deployment into the system, where the combination was retrieved after examining the model’s accuracy results across the local database. Secondly, we represent the outcomes of testing our surveillance car robot.

The testing set had 10% of the database and 243 data samples. By evaluating the model performance, we determined the optimal model, which consists of HC + VGG + KNN, has the highest accuracy rate of 99.00%. Starting with the confusion matrix given in Table 5.6, the samples labeled "Class_0" and "Class_1", were all correctly identified, whereas the class labeled "Class_2" had two miss-classified cases.

Table 4.5. Confusion matrix of the HC+VGG+ KNN model on local database

Predicted Class Actual Class	Class_0	Class_1	Class_2
Class_0	78	0	0
Class_1	0	83	0
Class_2	0	2	80

We can witness in Figure 4.8 the performance of the other models in terms of accuracy outcomes. Noting that the HC + ResNet + KNN model likewise yields a high accuracy of 98.76% with a slight deference to the chosen model.

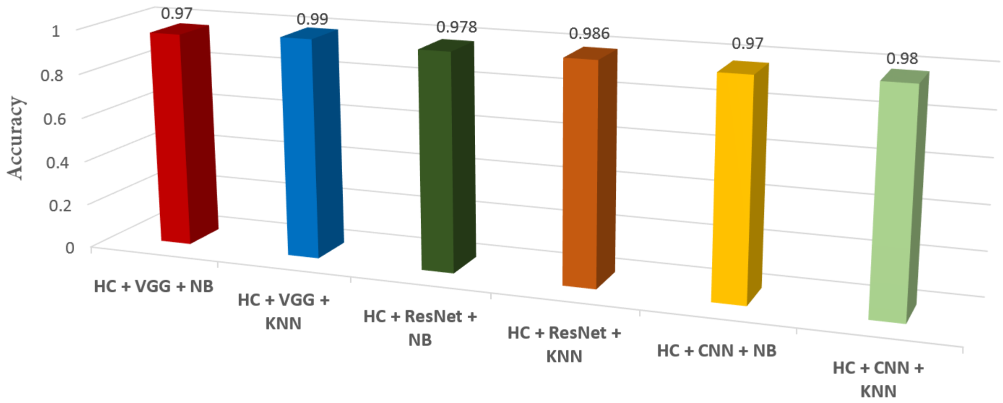


Fig. 4.8: Models performance over the local datataset [103].

The effectiveness of the surveillance car robot developed throughout the research has been evaluated in a practical context. The average response time of the system was 0.30 seconds in real time. The HC + VGG + KNN model, which has produced accurate predictions based on the collected face images, will assess and process the live video streaming provided by the robot. Figure 4.9 depicts a system reaction to a known instance where the face was effectively detected inside a domestic environment.

We altered the posture of the face in order to conduct more tests on the surveillance robot. Due to its non-frontal posture, as seen in Figure 4.10-a, the model was unable to distinguish the characteristics. Our edge robot was placed in a different setting outside the home during overcast weather, where it performed as it should in terms of movement, video streaming, and picture capture, but the processing procedures were hampered by the lighting, making the model unable to distinguish faces. Figure 4.10-b provides an illustration of the case scenario.

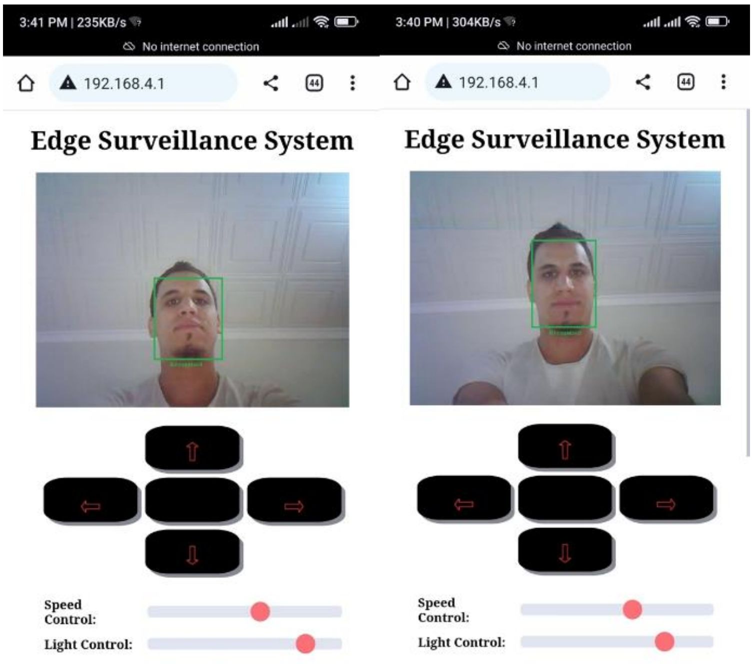
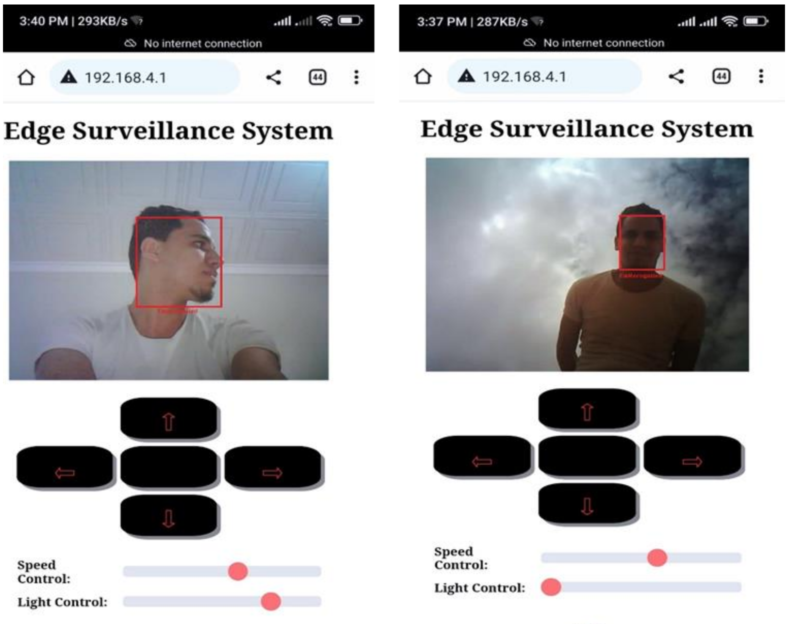


Fig. 4.9: Edge system response for normal case scenario [103].



(a) **(b)**

Fig. 4.10: (a) Edge robot response for a non-frontal face scenario.
(b) Edge robot response with bad lighting [103].

4.1.3 Conclusion

In this proposed scientific contribution, a FR system based on DL and ET was created. With the help of an established system that consists of numerous algorithms and a DL strategy, we have successfully created an edge car robot for SC monitoring. The system has a notice feature that sends the user an email when there are intruders or unknown individuals around. Our recommended solution offers a number of benefits, including being low-cost and inexpensive, consuming less power, being lightweight, supporting remote usage, and efficiently completing the task of real-time recognition. This research also demonstrates how complex AI models may now be executed on edge devices with specialized computational capabilities while preserving effectiveness and efficiency in terms of outputs. To give the model the best performance possible, additional work and more topics need to be covered, such as lightning changes since they have an impact on the outcomes of the predictions, face positioning, and facial emotions.

4.2 Face Recognition Surveillance System IoT and Deep Learning Based

Other contributions [103][104] were made with a proposed solution to the surveillance security aspect, a low-cost IoT surveillance systems using FR technology. The solutions suggested in these contributions was built using IoT ESP32 gadgets based on the DL high-efficacy CNN model. An Arduino-based solutions that has been implemented and built in a generalized way to be integrated into the CV field in form of smart applications that can be used in different domains and various industries, like door access control, light control, smart attendance systems, etc., and in verification or authentication services in general. It shows many advantages such affordability, high efficiency in terms of accuracy with a real-time response, lightweight solution with less power consumption, portability, and remote utilization.

4.2.1 Methodology

Figure 4.11 illustrates the proposed surveillance solution. As shown in the figure, the suggested FR system combines a CNN model for feature extraction and classification with an IoT device based on an ESP-32 CAM board already linked to the OV2640 camera module for live streaming and uses the HC algorithm for face detection. The ESP32-CAM module was configured on a local IP address to become a live streaming web server. Through those modules, the input video data will be processed using the OpenCV library. Each frame will be handled as an image to perform the preprocessing steps, from image reshaping, normalizing, face detection, feature extraction, and face recognition. An email alert notification will be sent to the system administrator, including the captured image as an attachment in the event of an unknown face. Additionally, our FR is available as a web application. The results are displayed on the screen as a drawn rectangle in green for recognized faces and red for the opposite scenario.

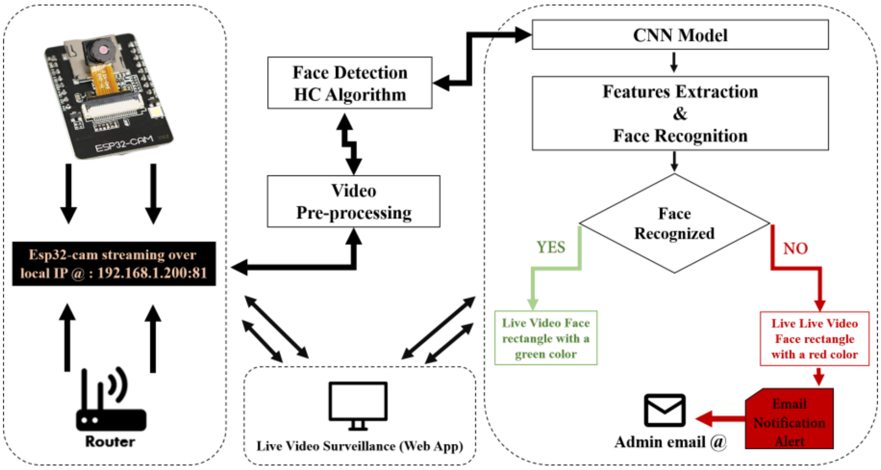

ESP32-CAM
Esp32-cam streaming over local IP @ : 192.168.1.200:81
Router
Face Detection HC Algorithm
Video Pre-processing
Live Video Surveillance (Web App)
CNN Model
Features Extraction & Face Recognition
Face Recognized
YES
NO
Live Video Face rectangle with a green color
Live Live Video Face rectangle with a red color
Email Notification Alert
Admin email @

Fig. 4.11: Proposed FR surveillance IoT system [103].

4.2.2 Experiments and Results

Many experiments were made to test and represent the performance of our FR system. Starting with the CNN model validation over four different benchmarks (the LFW, AR, ORL, and YALE databases). Figure 4.12 demonstrates the performance of the CNN model over these benchmarks and the achieved accuracy for each one. We report that the accuracy generally grows as the number of epochs likewise increases; until a certain number of epochs, it will either drop or remain the same. For all four databases, the optimum number of training epochs was between 160 and 200, with a little difference that didn't affect the results; therefore, we adjusted it to 200. The model obtains a better result on the YALE database with 98.95% accuracy and a low one in the LFW database with 95.09% accuracy, while an accuracy of 96.98% and 98.17% was noted on the AR and ORL databases, respectively.

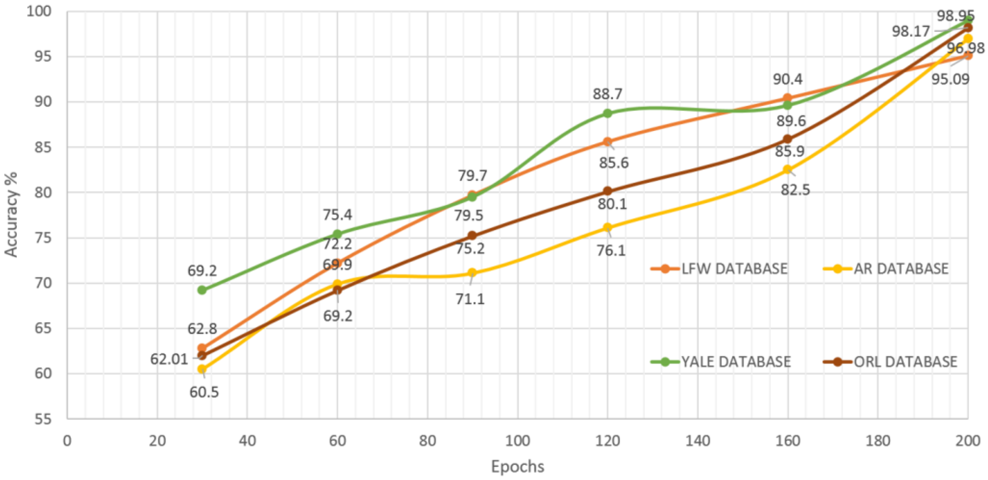


Fig. 4.12: CNN model validation results [103].

Tests were also made on a locally created database with five labeled classes. The model shows a high prediction efficiency, as can be seen in the confusion matrix represented in Table 5.7, we had only four miss-classified examples from the classes labeled “Class_2” and “Class_1”. While for the other labeled classes, all the samples were correctly classified. Figure 4.13 represents the class-wise accuracy, which illustrates the CNN model accuracy for each class. Finally, the accuracy and loss rates are recorded as 99.25% and 0.08, respectively.

Table 4.6. Confusion matrix of CNN model.

Predicted Class / Actual Class	Class_0	Class_1	Class_2	Class_3	Class_4
Class_0	81	0	0	0	0
Class_1	0	79	1	1	0
Class_2	0	0	79	1	1
Class_3	0	0	0	81	0
Class_4	0	0	0	0	81

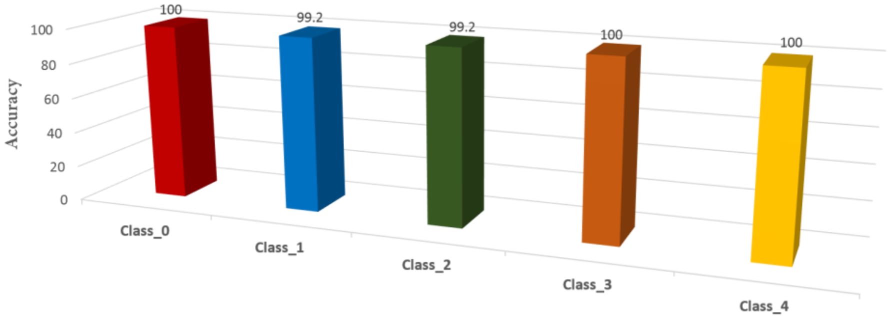


Fig. 4.13: Class wise accuracy [103].

4.2.3 Conclusion

In this work, a FR system using an IoT module based on DL was presented, where we successfully implemented a security system that uses HC algorithm and a CNN model. The suggested solution makes it easy for people to handle and integrate into their daily lives, with the possibility of using it from a distance, unlike other biometric solutions. The system has several advantages, such as being low-cost, less power-consuming, light-weight, and operating in real time with a user email notification. Further work and more aspects need to be covered in the future, such as testing this security system in a real-world application. Also, testing the system under various environmental conditions like lightning changes or in the face itself, taking into consideration face positioning or facial expressions.

4.3 Impact of our Contributions

The development of an IoT-based FR system represents a significant advancement in technology with far-reaching applications in various sectors. This proposed solution leverages the power of the IoT and advanced FR algorithms to provide enhanced security, efficiency, and convenience. The integration of these technologies enables real-time data processing and decision-making, making it a valuable asset in surveillance for SCs and improving our daily lives.

The provided solution can be used in the context of SCs specifically and in our daily life in general, the IoT-based FR system can play a crucial role in enhancing public safety and operational efficiency. Here are some key areas where it can be utilized:

1. **Real-Time Monitoring**: The system can be deployed across various public spaces, such as parks, transportation hubs, and city streets, to monitor activities in real-time. This capability allows for immediate identification of known criminals or persons of interest, helping law enforcement agencies to act swiftly and prevent potential crimes.
2. **Incident Response**: By providing instant alerts when suspicious activities are detected, the system enables quicker response times from emergency services. This proactive approach can significantly reduce the impact of criminal activities.
3. **Secure Access to Public Facilities**: Government buildings, airports, and other critical infrastructure can benefit from enhanced security measures provided by FR technology. Only authorized personnel can gain access, reducing the risk of unauthorized entry and potential threats.
4. **Automated Entry Systems**: In SC environments, FR can streamline access to various services. For instance, public transportation systems can use FR for ticketless travel, improving convenience for commuters.
5. **Identification of Traffic Violations**: The system can be integrated with traffic cameras to identify and record traffic violations, such as speeding or running red lights. This data can then be used to enforce traffic laws more effectively.
6. **Enhancing Traffic Flow**: By analyzing real-time data, city planners can optimize traffic flow and reduce congestion, contributing to a smoother and more efficient transportation network.

Beyond the scope of SCs, the IoT-based FR system can significantly impact our daily lives in various ways:

7. **Home Security**:
8. **Smart Home Integration**: By incorporating FR technology into smart home systems, homeowners can enhance the security of their residences. The system

can recognize family members and trusted visitors, granting them access while alerting the homeowner of any unidentified individuals.

9. **Automated Features**: The system can automate various home functions, such as lighting and climate control, based on the presence and preferences of recognized individuals.
10. **Employee Attendance and Access**: In corporate settings, FR can streamline employee attendance tracking and access control. This reduces the need for physical ID cards and improves overall security within the workplace.
11. **Visitor Management**: The system can also manage visitor access, ensuring that only authorized individuals can enter specific areas, thereby protecting sensitive information and assets.
12. **Personalized Shopping**: Retailers can use FR to enhance customer experiences by offering personalized recommendations and promotions based on the recognition of repeat customers.
13. **Loss Prevention**: The system can help in identifying known shoplifters, thereby reducing theft and enhancing overall store security.

Conclusion and Future Work

Nowadays, technology is changing our lives every day and swiftly, so that our society is governed by innovation, scientific advancement, and the applications of AI. Our research focuses on representing the power of technological solutions in the SCs surveillance and security aspects. Showing the advances that have been used and can be deployed in order to provide smart, effective solutions. Technologies like ET, IoT, CV, and DL are the leading solutions in the surveillance domain, which is what our work is mainly focused on. In this thesis, we began on a quest to study the convergence of modern technology and urban surveillance, delving into the domain of SC surveillance employing mainly IoT devices and FR systems. Our research aims to identify the implications, problems, and opportunities connected with this integration, offering insight into the shifting picture of urban security and citizen safety.

Finding, focusing, and proposing smart, leading technological surveillance solutions is a must nowadays because security problems and people's safety have become more crucial than ever since this technological progress wave led to massive data creation and exchange that put our personal information in danger of being spread and hacked very easily.

So, that lead to providing safety to individuals becomes a number one consideration and a must in today's life, such as in the case of this study about SC, where the goal is to raise inhabitants' overall quality of life in urban areas through the use of smart surveillance systems that integrate advanced technologies, including smart video camera devices, sensors, and data analytics techniques. Our study exposed various problems, ranging from privacy concerns to technological restrictions. It became obvious that striking a careful balance between security imperatives and individual privacy rights is vital. Moreover, the necessity for continual breakthroughs in ML algorithms such as CV solutions and DL models for problems like FR, paired with effective cybersecurity measures, has been identified as an essential area for future research. Additionally, the integration of AI for real-time threat detection and the creation of decentralized, secure data storage solutions are potential paths for further investigation. Through an in-depth review of the current literature, we discovered the multiple dynamics of SC, underscoring the vital role of IoT devices and FR technologies. We explored their synergistic potential for enhancing surveillance paradigms. Additionally, our study and case analyses underlined the necessity of privacy preservation, ethical issues, and public acceptance in the deployment of such monitoring technologies.

This thesis focuses on the development of a sophisticated and efficient smart solution tailored for the surveillance of smart cities. Drawing inspiration from the convergence of IoT ET, FR, and DL, the proposed system aims to revolutionize the traditional paradigms of urban surveillance. By harnessing the power of interconnected devices, advanced FR algorithms, and the learning capabilities of DNN, this research endeavors to create a holistic and adaptive surveillance framework. Through the amalgamation of these technologies, this thesis seeks to address the pressing need for a comprehensive and intelligent SC surveillance system. By providing a detailed exploration of the design, implementation, and evaluation of the proposed solution, this research endeavors to contribute to the growing body of knowledge in the field of smart city development and urban security. Ultimately, the aim is to establish a blueprint for the integration these technologies paving the way for safer, more secure, and resilient SCs of the future.

As we conclude this study, it is obvious that SC Surveillance, aided by IoT devices and FR technology, has the transformative capacity to reshape urban security. By embracing ethical principles, fostering openness, and including individuals in decision-making processes, we can pave the way for a secure, inclusive future. SCs, when coupled with reasonable smart surveillance techniques, have the ability to create places where safety and privacy coexist together, improving the lives of inhabitants and promoting a sense of belonging in the digital era.

Future work

More aspects need to be covered, and future work needs to be made, mainly the following:

- **Robustness and Security:** Develop strong algorithms and models to boost the accuracy of FR systems, especially in tough settings including low light, occlusions, and various face expressions. Investigate security techniques, including encryption and blockchain technology, to preserve the acquired surveillance data from unlawful access and manipulation.
- **Technological Advancements:** Analyze the potential integration of new technologies such as edge computing, 5G networks, and quantum computing to boost the efficiency and accuracy of IoT devices and FR systems. Study the improvements in DL approaches for FR, including generative adversarial networks and reinforcement learning, and their impact on SC surveillance.
- **Human-Computer Interaction:** Research human-computer interaction elements affecting public interfaces of SC surveillance systems, emphasizing user experience, accessibility, and user acceptance. Explore the design of user-friendly interfaces for citizens to access, regulate, and comprehend the data collected by surveillance systems.
- **Public-Private Partnerships:** Investigate relationships between public agencies, commercial firms, and academic organizations to stimulate innovation in SC surveillance technology.

Bibliography

[1] Xu, S., Wang, J., Shou, W., Ngo, T., Sadick, A.-M., & Wang, X. (2020). Computer Vision Techniques in Construction: A Critical Review. In Archives of Computational Methods in Engineering (Vol. 28, Issue 5, pp. 3383–3397). Springer Science and Business Media LLC. https://doi.org/10.1007/s11831-020-09504-3

[2] Szeliski, R. (2022). Computer Vision. In Texts in Computer Science. Springer International Publishing. https://doi.org/10.1007/978-3-030-34372-9

[3] Feng, X., Jiang, Y., Yang, X., Du, M., & Li, X. (2019). Computer vision algorithms and hardware implementations: A survey. In Integration (Vol. 69, pp. 309–320). Elsevier BV. https://doi.org/10.1016/j.vlsi.2019.07.005

[4] Al-Kaff, A., Martín, D., García, F., Escalera, A. de la, & María Armingol, J. (2018). Survey of computer vision algorithms and applications for unmanned aerial vehicles. In Expert Systems with Applications (Vol. 92, pp. 447–463). Elsevier BV. https://doi.org/10.1016/j.eswa.2017.09.033

[5] Bhatt, D., Patel, C., Talsania, H., Patel, J., Vaghela, R., Pandya, S., Modi, K., & Ghayvat, H. (2021). CNN Variants for Computer Vision: History, Architecture, Application, Challenges and Future Scope. In Electronics (Vol. 10, Issue 20, p. 2470). MDPI AG. https://doi.org/10.3390/electronics10202470

[6] Integration of Computer Vision and Natural Language Processing in Multimedia Robotics Application. (2022). In Information Sciences Letters (Vol. 11, Issue 3, pp. 765–775). Natural Sciences Publishing. https://doi.org/10.18576/isl/110309

[7] LeCun, Y.MNIST demos on Yann LeCun's website. Available at: http://yann.lecun.com/exdb/lenet/ (Accessed: 30 May 2023).

[8] Medical Imaging Specialist Melbourne: Melbourne Radiology (2023) Melbourne Radiology Clinic. Available at: https://www.melbourneradiology.com.au/ (Accessed: 01 December 2023).

[9] iTesLab. We put artificial intelligence at the service of human intelligence, iTeslab. Available at: https://www.iteslab.com/ (Accessed: 30 March 2023).

[10] Shetty, S. K., & Siddiqa, A. (2019). Deep Learning Algorithms and Applications in Computer Vision. In International Journal of Computer Sciences and Engineering (Vol. 7, Issue 7, pp. 195–201). ISROSET: International Scientific Research Organization for Science, Engineering and Technology. https://doi.org/10.26438/ijcse/v7i7.195201

[11] Zaidi, S. S. A., Ansari, M. S., Aslam, A., Kanwal, N., Asghar, M., & Lee, B. (2021). A Survey of Modern Deep Learning based Object Detection Models (Version 2). arXiv. https://doi.org/10.48550/ARXIV.2104.11892

[12] Kaur, J., & Singh, W. (2022). Tools, techniques, datasets and application areas for object detection in an image: a review. In Multimedia Tools and Applications (Vol. 81, Issue 27, pp. 38297–38351). Springer Science and Business Media LLC. https://doi.org/10.1007/s11042-022-13153-y

[13] Garau, N., & Conci, N. (2023). CapsulePose: A variational CapsNet for real-time end-to-end 3D human pose estimation. Neurocomputing, 523, 81–91. https://doi.org/10.1016/j.neucom.2022.11.097

[14] Porikli, F., & Yilmaz, A. (2012). Object Detection and Tracking. In Studies in Computational Intelligence (pp. 3–41). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-28598-1_1

[15] Borji, A., Cheng, M.-M., Hou, Q., Jiang, H., & Li, J. (2019). Salient object detection: A survey. In Computational Visual Media (Vol. 5, Issue 2, pp. 117–150). Springer Science and Business Media LLC. https://doi.org/10.1007/s41095-019-0149-9

[16] Belhadi, A., Djenouri, Y., Srivastava, G., Djenouri, D., Lin, J. C.-W., & Fortino, G. (2021). Deep learning for pedestrian collective behavior analysis in smart cities: A model of group trajectory outlier detection. In Information Fusion (Vol. 65, pp. 13–20). Elsevier BV. https://doi.org/10.1016/j.inffus.2020.08.003

[17] Nayak, R., Pati, U. C., & Das, S. K. (2021). A comprehensive review on deep learning-based methods for video anomaly detection. In Image and Vision Computing (Vol. 106, p. 104078). Elsevier BV. https://doi.org/10.1016/j.imavis.2020.104078

[18] Kumaran, S. K., Dogra, D. P., & Roy, P. P. (2019). Anomaly Detection in Road Traffic Using Visual Surveillance: A Survey. arXiv. https://doi.org/10.48550/ARXIV.1901.08292

[19] Kumaran, S. K., Dogra, D. P., & Roy, P. P. (2019). Anomaly Detection in Road Traffic Using Visual Surveillance: A Survey. arXiv. https://doi.org/10.48550/ARXIV.1901.08292

[20] Khan, A. A., Nauman, M. A., Shoaib, M., Jahangir, R., Alroobaea, R., Alsafyani, M., Binmahfoudh, A., & Wechtaisong, C. (2022). Crowd Anomaly Detection in Video Frames Using Fine-Tuned AlexNet Model. In Electronics (Vol. 11, Issue 19, p. 3105). MDPI AG. https://doi.org/10.3390/electronics11193105

[21] Patrikar, D. R., & Parate, M. R. (2022). Anomaly detection using edge computing in video surveillance system: review. In International Journal of Multimedia Information Retrieval (Vol. 11, Issue 2, pp. 85–110). Springer Science and Business Media LLC. https://doi.org/10.1007/s13735-022-00227-8

[22] Silva, B. N., Khan, M., & Han, K. (2018). Towards sustainable smart cities: A review of trends, architectures, components, and open challenges in smart cities. In Sustainable Cities and Society (Vol. 38, pp. 697–713). Elsevier BV. https://doi.org/10.1016/j.scs.2018.01.053

[23] Lai, C. S., Jia, Y., Dong, Z., Wang, D., Tao, Y., Lai, Q. H., Wong, R. T. K., Zobaa, A. F., Wu, R., & Lai, L. L. (2020). A Review of Technical Standards for Smart Cities. In Clean Technologies (Vol. 2, Issue 3, pp. 290–310). MDPI AG. https://doi.org/10.3390/cleantechnol2030019

[24] Khatoun, R., & Zeadally, S. (2016). Smart cities. In Communications of the ACM (Vol. 59, Issue 8, pp. 46–57). Association for Computing Machinery (ACM). https://doi.org/10.1145/2858789

[25] Arasteh, H., Hosseinnezhad, V., Loia, V., Tommasetti, A., Troisi, O., Shafie-khah, M., & Siano, P. (2016). Iot-based smart cities: A survey. In 2016 IEEE 16th International Conference on Environment and Electrical Engineering (EEEIC). 2016

IEEE 16th International Conference on Environment and Electrical Engineering (EEEIC). IEEE. https://doi.org/10.1109/eeeic.2016.7555867

[26] Crnjac, M., Veža, I., & Banduka, N. (2017). From Concept to the Introduction of Industry 4.0. In International Journal of Industrial Engineering and Management (Vol. 8, Issue 1, pp. 21–30). Faculty of Technical Sciences. https://doi.org/10.24867/ijiem-2017-1-103

[27] Conti, M., Passarella, A., & Das, S. K. (2017). The Internet of People (IoP): A new wave in pervasive mobile computing. In Pervasive and Mobile Computing (Vol. 41, pp. 1–27). Elsevier BV. https://doi.org/10.1016/j.pmcj.2017.07.009

[28] Doku, R., & Rawat, D. B. (2019). Big Data in Cybersecurity for Smart City Applications. In Smart Cities Cybersecurity and Privacy (pp. 103–112). Elsevier. https://doi.org/10.1016/b978-0-12-815032-0.00008-1

[29] Al-Turjman, F., & Lemayian, J. P. (2020). Intelligence, security, and vehicular sensor networks in internet of things (IoT)-enabled smart-cities: An overview. In Computers &amp; Electrical Engineering (Vol. 87, p. 106776). Elsevier BV. https://doi.org/10.1016/j.compeleceng.2020.106776

[30] Javed, A. R., Ahmed, W., Pandya, S., Maddikunta, P. K. R., Alazab, M., & Gadekallu, T. R. (2023). A Survey of Explainable Artificial Intelligence for Smart Cities. In Electronics (Vol. 12, Issue 4, p. 1020). MDPI AG. https://doi.org/10.3390/electronics12041020

[31] Guo, Y.-M., Huang, Z.-L., Guo, J., Li, H., Guo, X.-R., & Nkeli, M. J. (2019). Bibliometric Analysis on Smart Cities Research. In Sustainability (Vol. 11, Issue 13, p. 3606). MDPI AG. https://doi.org/10.3390/su11133606

[32] Toh, C. K. (2020). Security for smart cities. In IET Smart Cities (Vol. 2, Issue 2, pp. 95–104). Institution of Engineering and Technology (IET). https://doi.org/10.1049/iet-smc.2020.0001

[33] Al-Turjman, F., Zahmatkesh, H., & Shahroze, R. (2019). An overview of security and privacy in smart cities' IoT communications. In Transactions on Emerging Telecommunications Technologies (Vol. 33, Issue 3). Wiley. https://doi.org/10.1002/ett.3677

[34] Szum, K. (2021). IoT-based smart cities: a bibliometric analysis and literature review. In Engineering Management in Production and Services (Vol. 13, Issue 2, pp. 115–136). Walter de Gruyter GmbH. https://doi.org/10.2478/emj-2021-0017

[35] Ghazal, T. M., Hasan, M. K., Alshurideh, M. T., Alzoubi, H. M., Ahmad, M., Akbar, S. S., Al Kurdi, B., & Akour, I. A. (2021). IoT for Smart Cities: Machine Learning Approaches in Smart Healthcare—A Review. In Future Internet (Vol. 13, Issue 8, p. 218). MDPI AG. https://doi.org/10.3390/fi13080218

[36] Hassankhani, M., Alidadi, M., Sharifi, A., & Azhdari, A. (2021). Smart City and Crisis Management: Lessons for the COVID-19 Pandemic. In International Journal of Environmental Research and Public Health (Vol. 18, Issue 15, p. 7736). MDPI AG. https://doi.org/10.3390/ijerph18157736

[37] Sharifi, A., Khavarian-Garmsir, A. R., & Kummitha, R. K. R. (2021). Contributions of Smart City Solutions and Technologies to Resilience against the COVID-19 Pandemic: A Literature Review. In Sustainability (Vol. 13, Issue 14, p. 8018). MDPI AG. https://doi.org/10.3390/su13148018

[38] Kylili, A., Afxentiou, N., Georgiou, L., Panteli, C., Morsink-Georgalli, P.-Z., Panayidou, A., Papouis, C., & Fokaides, P. A. (2020). The role of Remote Working in smart cities: lessons learnt from COVID-19 pandemic. In Energy Sources, Part A: Recovery, Utilization, and Environmental Effects (pp. 1–16). Informa UK Limited. https://doi.org/10.1080/15567036.2020.1831108

[39] Zhou, Z., Chen, X., Li, E., Zeng, L., Luo, K., & Zhang, J. (2019). Edge Intelligence: Paving the Last Mile of Artificial Intelligence With Edge Computing. In Proceedings of the IEEE (Vol. 107, Issue 8, pp. 1738–1762). Institute of Electrical and Electronics Engineers (IEEE). https://doi.org/10.1109/jproc.2019.2918951

[40] Guevara, L., & Auat Cheein, F. (2020). The Role of 5G Technologies: Challenges in Smart Cities and Intelligent Transportation Systems. In Sustainability (Vol. 12, Issue 16, p. 6469). MDPI AG. https://doi.org/10.3390/su12166469

[41] Ullah, Z., Al-Turjman, F., Mostarda, L., & Gagliardi, R. (2020). Applications of Artificial Intelligence and Machine learning in smart cities. In Computer Communications (Vol. 154, pp. 313–323). Elsevier BV. https://doi.org/10.1016/j.comcom.2020.02.069

[42] Gartner_Inc. Understanding Gartner's hype cycles. Gartner. https://www.gartner.com/en/documents/3887767. Accessed : 09-18-2023

[43] Wang, X., Han, Y., Leung, V. C. M., Niyato, D., Yan, X., & Chen, X. (2019). Convergence of Edge Computing and Deep Learning: A Comprehensive Survey. arXiv. https://doi.org/10.48550/ARXIV.1907.08349

[44] Laufs, J., Borrion, H., & Bradford, B. (2020). Security and the smart city: A systematic review. In Sustainable Cities and Society (Vol. 55, p. 102023). Elsevier BV. https://doi.org/10.1016/j.scs.2020.102023

[45] Haque, A. K. M. B., Bhushan, B., & Dhiman, G. (2021). Conceptualizing smart city applications: Requirements, architecture, security issues, and emerging trends. In Expert Systems (Vol. 39, Issue 5). Wiley. https://doi.org/10.1111/exsy.12753

[46] Swathi, K., Sandeep, T. U., & Ramani, A. R. (2018). Performance Analysis of Microcontrollers Used In IoT Technology. International journal of scientific research in science, engineering and technology, 4(4), 1268-1273.

[47] Cheour, R., Khriji, S., abid, M., & Kanoun, O. (2020). Microcontrollers for IoT: Optimizations, Computing Paradigms, and Future Directions. In 2020 IEEE 6th World Forum on Internet of Things (WF-IoT). 2020 IEEE 6th World Forum on Internet of Things (WF-IoT). IEEE. https://doi.org/10.1109/wf-iot48130.2020.9221219

[48] Raspberrypi. https://www.raspberrypi.org/. Accessed : 09-05-2023

[49] Arduino. https://www.arduino.cc/ . Accessed : 09-05-2023

[50] Gartner_Inc. http://esp32.net/. Accessed : 09-05-2023

[51] Espressif. https://www.espressif.com/en/products/socs/esp8266. Accessed : 09-05-2023

[52] Jetson-Nano. https://developer.nvidia.com/embedded/jetson-nano-developer-kit. Accessed : 09-05-2023

[53] Beagleboard. https://www.beagleboard.org/boards/beaglebone-black. Accessed : 09-05-2023

[54] Particle. https://docs.particle.io/electron/. Accessed : 09-05-2023

[55] Hardkernel. https://www.hardkernel.com/. Accessed : 09-05-2023

[56] Helium. https://www.helium.com/ecosystem. Accessed : 09-05-2023

[57] Microchip. https://www.microchip.com/en-us/products/microcontrollers-and-microprocessors/8-bit-mcus. Accessed : 09-05-2023

[58] Viso. https://viso.ai/applications/computer-vision-applications-in-surveillance-and-security/. Accessed : 09-05-2023

[59] Kortli, Y., Jridi, M., Al Falou, A., & Atri, M. (2020). Face Recognition Systems: A Survey. In Sensors (Vol. 20, Issue 2, p. 342). MDPI AG. https://doi.org/10.3390/s20020342

[60] Adjabi, I., Ouahabi, A., Benzaoui, A., & Taleb-Ahmed, A. (2020). Past, Present, and Future of Face Recognition: A Review. In Electronics (Vol. 9, Issue 8, p. 1188). MDPI AG. https://doi.org/10.3390/electronics9081188

[61] Kaur, P., Krishan, K., Sharma, S. K., & Kanchan, T. (2020). Facial-recognition algorithms: A literature review. In Medicine, Science and the Law (Vol. 60, Issue 2, pp. 131–139). SAGE Publications. https://doi.org/10.1177/0025802419893168

[62] Bledsoe, W.W. (1964). The model method in facial recognition. Technical Report, Panoramic Research, Inc., Palo Alto, California.

[63] Goldstein, A. J., Harmon, L. D., & Lesk, A. B. (1971). Identification of human faces. In Proceedings of the IEEE (Vol. 59, Issue 5, pp. 748–760). Institute of Electrical and Electronics Engineers (IEEE). https://doi.org/10.1109/proc.1971.8254

[64] Turk, M., & Pentland, A. (1991). Eigenfaces for Recognition. In Journal of Cognitive Neuroscience (Vol. 3, Issue 1, pp. 71–86). MIT Press - Journals. https://doi.org/10.1162/jocn.1991.3.1.71

[65] Belhumeur, P.N., Hespanha, J.P. and Kriegman, D. (1997) Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 19, 711-720. http://dx.doi.org/10.1109/34.598228

[66] P. Viola, M. J. Jones, "Robust Real-Time Face Detection," International Journal of Computer Vision, vol. 57, pp. 137–154, 5 2004, https://doi.org/10.1023/B:VISI.0000013087.49260.fb.

[67] Fischer, A., & Igel, C. (2012). An introduction to restricted Boltzmann machines. In Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications: 17th Iberoamerican Congress, CIARP 2012, Buenos Aires, Argentina, September 3-6, 2012. Proceedings 17 (pp. 14-36). Springer Berlin Heidelberg.

[68] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems, 25.

[69] Schroff, F., Kalenichenko, D., & Philbin, J. (2015). FaceNet: A Unified Embedding for Face Recognition and Clustering. arXiv. https://doi.org/10.48550/ARXIV.1503.03832

[70] Taigman, Y., Yang, M., Ranzato, M., & Wolf, L. (2014). DeepFace: Closing the Gap to Human-Level Performance in Face Verification. In 2014 IEEE Conference on Computer Vision and Pattern Recognition. 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE. https://doi.org/10.1109/cvpr.2014.220

[71] Parkhi, O., Vedaldi, A., & Zisserman, A. (2015). Deep face recognition. In BMVC 2015-Proceedings of the British Machine Vision Conference 2015. British Machine Vision Association.

[72] Ouyang, W., Luo, P., Zeng, X., Qiu, S., Tian, Y., Li, H., Yang, S., Wang, Z., Xiong, Y., Qian, C., Zhu, Z., Wang, R., Loy, C.-C., Wang, X., & Tang, X. (2014). DeepID-Net: multi-stage and deformable deep convolutional neural networks for object detection (Version 1). arXiv. https://doi.org/10.48550/ARXIV.1409.3505

[73] Deng, J., Guo, J., Yang, J., Xue, N., Kotsia, I., & Zafeiriou, S. (2018). ArcFace: Additive Angular Margin Loss for Deep Face Recognition. arXiv. https://doi.org/10.48550/ARXIV.1801.07698

[74] Zhou, S., Xiao, S. 3D face recognition: a survey. Hum. Cent. Comput. Inf. Sci. 8, 35 (2018). https://doi.org/10.1186/s13673-018-0157-2

[75] Davis E. King. Dlib-ml: A Machine Learning Toolkit. Journal of Machine Learning Research 10, pp. 1755-1758, 2009

[76] Baltrušaitis, T., Robinson, P., & Morency, L. P. (2016, March). Openface: an open source facial behavior analysis toolkit. In 2016 IEEE winter conference on applications of computer vision (WACV) (pp. 1-10). IEEE.

[77] Firouzi, M., Fadaei, S., & Rashno, A. (2022). A New Framework for Canny Edge Detector in Hexagonal Lattice. In International Journal of Engineering (Vol. 35, Issue 8, pp. 1588–1593). International Digital Organization for Scientific Information (IDOSI). https://doi.org/10.5829/ije.2022.35.08b.15

[78] Stone, J. V. (2004). Independent component analysis: a tutorial introduction.

[79] Ahonen, T., Hadid, A., Pietikäinen, M. Face description with local binary patterns: Application to face recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2006, 28(12), 2037-2041.

[80] The ORL Database of Faces. http:// www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html. Accessed : 09-05-2023

[81] FERET. https://www.nist.gov/programs-projects/face-recognition-technology-feret. Accessed : 09-05-2023

[82] The AR Face Database. https://www2.ece.ohio-state.edu/~aleix/ARdatabase.html. Accessed : 09-05-2023

[83] The LFW Database (Labeled Faces in the Wild). http://vis-www.cs.umass.edu/lfw/. Accessed : 09-05-2023

[84] CASIA Web Face. http://www.cbsr.ia.ac.cn/english/CASIA-WebFace-Database.html. Accessed : 09-05-2023

[85] Elharrouss, O.; Almaadeed, N.; Al-Maadeed, S. LFR face dataset: Left-Front-Right dataset for pose-invariant face recognition in the wild. In Proceedings of the 2020 IEEE International Conference on Informatics, IoT, and Enabling Technologies (ICIoT), Doha, Qatar, 2020; 124-130.

[86] Wang, Z., Wang, G., Huang, B., Xiong, Z., Hong, Q., Wu, H., Yi, P., Jiang, K., Wang, N., Pei, Y., Chen, H., Miao, Y., Huang, Z., & Liang, J. (2020). Masked Face Recognition Dataset and Application (Version 2). arXiv. https://doi.org/10.48550/ARXIV.2003.09093

[87] Masud, M., Muhammad, G., Alhumyani, H., Alshamrani, S. S., Cheikhrouhou, O., Ibrahim, S., & Hossain, M. S. (2020). Deep learning-based intelligent face

recognition in IoT-cloud environment. In Computer Communications (Vol. 152, pp. 215–222). Elsevier BV. https://doi.org/10.1016/j.comcom.2020.01.050

[88] Faisal, F., & Hossain, S. A. (2019). Smart Security System Using Face Recognition on Raspberry Pi. In 2019 13th International Conference on Software, Knowledge, Information Management and Applications (SKIMA). 2019 13th International Conference on Software, Knowledge, Information Management and Applications (SKIMA). IEEE. https://doi.org/10.1109/skima47702.2019.8982466

[89] Mohi Uddin, K. M., Afrin Shahela, S., Rahman, N., Mostafiz, R., & Rahman, Md. M. (2022). Smart Home Security Using Facial Authentication and Mobile Application. In International Journal of Wireless and Microwave Technologies (Vol. 12, Issue 2, pp. 40–50). MECS Publisher. https://doi.org/10.5815/ijwmt.2022.02.04

[90] Kumar, T. A., Rajmohan, R., Pavithra, M., Ajagbe, S. A., Hodhod, R., & Gaber, T. (2022). Automatic Face Mask Detection System in Public Transportation in Smart Cities Using IoT and Deep Learning. In Electronics (Vol. 11, Issue 6, p. 904). MDPI AG. https://doi.org/10.3390/electronics11060904

[91] Jiménez-Bravo, D. M., Lozano Murciego, Á., Sales Mendes, A., Silva, L. A., & De La Iglesia, D. H. (2022). Edge Face Recognition System Based on One-Shot Augmented Learning. In International Journal of Interactive Multimedia and Artificial Intelligence (Vol. 7, Issue 6, p. 31). Universidad Internacional de La Rioja. https://doi.org/10.9781/ijimai.2022.09.001

[92] Malve, S., & Morade, S. S. (2021). Face Recognition Technology based Smart Doorbell System using Python's OpenCV library. INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT), 10(06).

[93] Yang, S., Wen, Y., He, L., Zhou, M., & Abusorrah, A. (2021). Sparse Individual Low-Rank Component Representation for Face Recognition in the IoT-Based System. In IEEE Internet of Things Journal (Vol. 8, Issue 24, pp. 17320–17332). Institute of Electrical and Electronics Engineers (IEEE). https://doi.org/10.1109/jiot.2021.3080084

[94] Irjanto, N. S., & Surantha, N. (2020). Home Security System with Face Recognition based on Convolutional Neural Network. In International Journal of Advanced Computer Science and Applications (Vol. 11, Issue 11). The Science and Information Organization. https://doi.org/10.14569/ijacsa.2020.0111152

[95] Hussain, T., Hussain, D., Hussain, I., AlSalman, H., Hussain, S., Ullah, S. S., & Al-Hadhrami, S. (2022). Internet of Things with Deep Learning-Based Face Recognition Approach for Authentication in Control Medical Systems. In S. Ahmad (Ed.), Computational and Mathematical Methods in Medicine (Vol. 2022, pp. 1–17). Hindawi Limited. https://doi.org/10.1155/2022/5137513

[96] Bhatia, P., Rajput, S., Pathak, S., & Prasad, S. (2018). IOT based facial recognition system for home security using LBPH algorithm. In 2018 3rd International Conference on Inventive Computation Technologies (ICICT). 2018 3rd International Conference on Inventive Computation Technologies (ICICT). IEEE. https://doi.org/10.1109/icict43934.2018.9034420

[97] Shaheen, M. H., Wani, M. Y., Raza, M. U., Tahir, A., & Nadeem, Y. (2021). Smart Intelligence Mirror System. Journal of Information Technology and Computing, 2(1), 30-40.

[98] R., M., Y., R., R., R., & A., S. (2020). Smart Home Security System using Iot, Face Recognition and Raspberry Pi. In International Journal of Computer Applications (Vol. 176, Issue 13, pp. 45–47). Foundation of Computer Science. https://doi.org/10.5120/ijca2020920105

[99] Kumar, M., Raju, K. S., Kumar, D., Goyal, N., Verma, S., & Singh, A. (2021). An efficient framework using visual recognition for IoT based smart city surveillance. In Multimedia Tools and Applications (Vol. 80, Issue 20, pp. 31277–31295). Springer Science and Business Media LLC. https://doi.org/10.1007/s11042-020-10471-x

[100] Uddin, K. M. M., Dey, S. K., Parvez, G. U., Mukta, A. S., & Acharjee, U. K. (2021). MirrorME: implementation of an IoT based smart mirror through facial recognition and personalized information recommendation algorithm. In International Journal of Information Technology (Vol. 13, Issue 6, pp. 2313–2322). Springer Science and Business Media LLC. https://doi.org/10.1007/s41870-021-00801-z

[101] Rahim, A., Zhong, Y., Ahmad, T., Ahmad, S., Pławiak, P., & Hammad, M. (2023). Enhancing Smart Home Security: Anomaly Detection and Face Recognition in Smart Home IoT Devices Using Logit-Boosted CNN Models. In Sensors (Vol. 23, Issue 15, p. 6979). MDPI AG. https://doi.org/10.3390/s23156979

[102] Medjdoubi, A., Meddeber, M., & Yahyaoui, K. (2024). Smart City Surveillance: Edge Technology Face Recognition Robot Deep Learning Based. In International Journal of Engineering (Vol. 37, Issue 1, pp. 25–36). International Digital Organization for Scientific Information (IDOSI). https://doi.org/10.5829/ije.2024.37.01a.03

[103] Medjdoubi, A., Meddeber, M., & Yahyaoui, K. (2022). Deep Learning IoT Based Smart Face Recognition System. In International Conference on Artificial Intelligence and Soft Computing (ICAISC). ITRESEARCH.

[104] Medjdoubi, A., Meddeber, M., & Yahyaoui, K. (2023). Face Recognition Surveillance System IoT and Deep Learning Based. International Conference on Engineering & Technology (ICET-23). ITAR.