Securing Teleoperated Robot: Classifying Human Operator Identity and Emotion through Motion-Controlled Robotic Behaviors

Xinyu Qi

Submitted in fulfilment of the requirements for the
Degree of Doctor of Philosophy

School of Engineering
College of Science and Engineering
University of Glasgow

March 2024

Abstract

Teleoperated robotic systems allow human operators to control robots from a distance, which mitigates the constraints of physical distance between the operators and offers invaluable applications in the real world. However, the security of these systems is a critical concern. System attacks and the potential impact of operators' inappropriate emotions can result in misbehavior of the remote robots, which poses risks to the remote environment. These concerns become particularly serious when performing mission-critical tasks, such as nuclear cleaning. This thesis explored innovative security methods for the teleoperated robotic system.

Common methods of security that can be used for teleoperated robots include encryption, robot misbehavior detection and user authentication. However, they have limitations for teleoperated robot systems. Encryption adds communication overheads to the systems. Robot misbehavior detection can only detect unusual signals on robot devices. The user authentication method secured the system primarily at the access point. To address this, we built motion-controlled robot platforms that allow for robot teleoperation and proposed methods of performing user classification directly on remote-controlled robotic behavioral data to enhance security integrity throughout the operation. We discussed in Chapter 3 and conducted 4 experiments. Experiments 1 and 2 demonstrated the effectiveness of our approach, achieving user classification accuracy of 95% and 93% on two platforms respectively, using motion-controlled robotic end-effector trajectories. The results in experiment 3 further indicated that control system performance directly impacts user classification efficacy. Additionally, we deployed an AI agent to protect user biometric identities, ensuring the robot's actions do not compromise user privacy in the remote environment in experiment 4. This chapter provided a foundation of methodology and experiment design for the next work.

Additionally, Operators' emotions could pose a security threat to the robot system. A remote robot operator's emotions can significantly impact the resulting robot's motions leading to unexpected consequences, even when the user follows protocol and performs permitted tasks. The recognition of a user operator's emotions in remote robot control scenarios is, however, under-explored. Emotion signals mainly are physiological signals, semantic information, facial expressions and bodily movements. However, most physiological signals are electrical signals and are vulnerable to motion artifacts, which can not acquire the accurate signal and is not suitable for teleoperated robot systems. Semantic information and facial expressions are some-

times not accessible and involve high privacy issues and add additional sensors to the teleoperated systems. We proposed the methods of emotion recognition through the motion-controlled robotic behaviors in Chapter 4. This work demonstrated for the first time that the motion-controlled robotic arm can inherit human operators' emotions and emotions can be classified through robotic end-effector trajectories, achieving an 83.3% accuracy. We developed two emotion recognition algorithms using Dynamic Time Warping (DTW) and Convolutional Neural Network (CNN), deriving unique emotional features from the avatar's end-effector motions and joint spatial-temporal characteristics. Additionally, we demonstrated through direct comparison that our approach is more appropriate for motion-based telerobotic applications than traditional ECG-based methods. Furthermore, we discussed the implications of this system on prominent current and future remote robot operations and emotional robotic contexts.

By integrating user classification and emotion recognition into teleoperated robotic systems, this thesis lays the groundwork for a new security paradigm that enhances both the safety of remote operations. Recognizing users and their emotions allows for more contextually appropriate robot responses, potentially preventing harm and improving the overall quality of teleoperated interactions. These advancements contribute significantly to the development of more adaptive, intuitive, and human-centered HRI applications, setting a precedent for future research in the field.

Contents

List of Tables

List of Figures

Acknowledgements

I would like to express my sincere gratitude to my supervisors, Philip G. Zhao and Imran Muhammad, for opening the door to this opportunity.

My heartfelt gratitude goes to my family, whose unwavering support and encouragement have been my constant source of strength.

I am also grateful to my friends who have been with me throughout this journey, offering their help and always being there to listen throughout this adventure.

My sincere thanks also go to my colleagues for their invaluable wisdom, empathy, and support.

To everyone who has been a part of this journey, thank you for inspiring me to face the unknown with bravery and challenge with confidence.

Declaration

I certify that all work presented in this thesis for a PhD degree from the University of Glasgow was entirely carried out by the author.

I declare that this thesis has been produced in accordance with the University of Glasgow's Code of Good Practice in Research.

Chapter 1

Introduction

1.1 Motivation

Robot teleoperation is that a human operator remotely controls the robotic system to interact with the remote environments [1]. The robot acts as the slave executing tasks under the direction of the human operator as the master. The teleoperation methods mainly include direct control using a controller (e.g. joysticks) and multimodal teleoperation control (e.g. human motion). Teleoperated robots bridge the physical gap between humans and remote environments, regarded as a key application within the field of Human-Robot Interaction (HRI). Their importance was highlighted during the COVID-19 pandemic, demonstrating their significant role across various scenarios. Firstly, they enable human to execute tasks in hazardous or inaccessible environments. For instance, teleoperated robots can be deployed to carry out disinfection in epidemic areas and cleaning in nuclear power stations [2], thereby minimizing the human risk of secondary exposure to dangerous environments. Additionally, teleoperated robotics is beneficial for individuals with disabilities. For example, wheelchair users could perform tasks requiring upper body or hand movement remotely. The teleoperated robots can be used for remote health as well. For example, a remote full body motion controlled robot can be used for remote rehabilitation [3] and telesurgery [4]. Moreover, teleoperated robotics can be applied to remote education. For example, it can present in the class as a substitute for teachers to mitigate the students' potential pressure [5].

However, ensuring robotic security becomes a crucial concern, especially when these vital applications and missions are involved [6]. Common security methods, including encryption, robot misbehavior detection and user authentication, that can be used for enhancing the security of remote control robotic systems exist limitations. Encryption is a security method for signal transmission by encoding and decoding the signals, but it adds overheads on communication and decreases the control efficiency. The robot misbehavior detection method is to detect the unusual signals of robot devices, but it is applied to an autonomous robot and can not protect the remote control robot system, especially for impersonation attacks. For the user authentication method,

there are three main categorizations including knowledge-based, token-based and biometrics-based. The knowledge-based method [7] requires users to remember predefined passwords and use them to log in to the system, but the system is vulnerable to attacks by any individuals who know the passwords. The token-based method allows users a preallocate tokens to take, but the tokens are burdensome and are easy to theft. The biometric-based method requires users' physical or behavioral biometric information. But the traditional physical biometrical information is easy to be theft. For example, fingerprints can be reproduced from a photograph and facial images can be downloaded from users' social media. Behavioral biometric user authentication that allows continuous authentication is a relatively new concept requiring users' body gestures [8], but existing methods were applied to the user access point and can not the system security after user login. The combination of the single modality of user authentication is Multi-Factor Authentication (MFA), but it has a similar limitation to behavioral biometric user authentication, which can not protect the whole robot system's security. These methods primarily secure the human-robot communication at the point of access, however, they may not address the vulnerability of command signals being tampered with during transmission over long distances.

To mitigate this risk, employing user authentication directly through the robotic data offers a more robust solution, which can protect the whole system's security without decreasing the controlling efficiency. We proposed a user classification method on robotic data to verify the legitimacy of the commands, ensuring the integrity of system security. This approach not only reinforces the security framework but also maintains the trustworthiness and reliability of the remote-controlled robotic system. We built a motion-controlled robotic system that allows for robot teleoperation and analysed motion-controlled robotic behaviors. Instead of relying on data from human operators, we applied user classification techniques directly to the robotic data. Furthermore, we demonstrated the adaptability of our method across two distinct motion-controlled robotic platforms, each utilizing different motion capture devices and types of robots. After the system identifies users, it can provide personalization, such as specific needs and preferences, for different users. For example, users have different preferences on robotic control velocity, influencing how comfortable and suitable they feel while managing the robot. Thus, we performed user classification under different controlling parameters. Besides, a concern with direct user classification on robots is the potential risk of compromising operator privacy, especially when data is transmitted to remote locations. To address this privacy challenge, we introduced a reinforcement learning method that the robot can execute commands that do not contain user biometric identities, which safeguards users' privacy effectively.

The way to express emotions is different among different human. Thus, human operators' emotions, which is an indispensable factor in HRI, should be considered after classifying the users. Emotion influences how messages are sent, received, and interpreted [9]. The ability to regulate, express and interpret emotions is emotion intelligence, which plays an important role in HRI and is considered multidisciplinary encompassing psychology, social sciences, cogni-

tive science, artificial intelligence, design, engineering, and computer science [10]. Endowing robotics with emotional intelligence allows robots can interpret human emotions and give feedback accordingly, thus having the ability to understand human messages appropriately, maintain relationships and enhance the decision-making process. It can help to build human-centred HRI. Firstly, in the field of customer service, understanding human emotions can enhance the effectiveness of services by providing personalized services [11]. Secondly, in the field of industry, understanding workers' emotions and giving empathy increase user satisfaction and acceptance of robots in working environments [12]. Thirdly, in healthcare settings, companion robots that have emotional intelligence for the elderly can provide comfort and companionship, potentially reducing feelings of loneliness [13]. Fourthly, in the field of education, robots can adapt their teaching methods to fit the child's emotional feedback. Lastly, in situations where robots assist humans in disaster zones, robots that recognize the stress [14] of human operators can avoid secondary damage to the local area.

Human emotions have a vital influence on teleoperated robot systems, however, they are rarely discussed in this field. Inappropriate emotions for specific scenarios can compromise system safety. For example, a driver's fatigue could lead to distraction while operating teleoperated vehicles, potentially causing severe accidents. More critically, if a surgeon is irritable during telesurgery, it could result in harm to an anesthetized patient who is unaware of the situation. This highlights the need for integrating emotional intelligence into teleoperated systems to assess the emotional states of operators, ensuring safer and more effective interactions. Thus, endowing robots with emotional intelligence is quite significant. For example, in the remote driving scenario, automated driving can be switched on when the driver's exhausted emotion is detected, which can lead to improved user satisfaction and greater efficiency in task completion.

A key technology for recognising emotion is emotion recognition (ER), which is a rapidly evolving interdisciplinary field including AI, natural language processing (NLP), cognitive sciences, social sciences and psychology [15]. Together, these disciplines collaborate to develop an intelligent system that is capable of sensing and interpreting the emotional states of humans. It aims to enhance human-computer interaction and create more empathetic and intuitive technological solutions. There are two main techniques for ER including sensing emotion signals and implementing AI to classify emotions. The sensing signals from human behaviors include physiological signals (e.g. heart signals), semantic signals (e.g. Twitter and Instagram), facial signals and body signals (e.g. shaking hands).

However, ER in the context of HRI remains less explored [16] compared with the human-computer interaction (HCI) context. Research of ER in HRI mainly focuses on a particular HRI application area: human-robot social interaction. Specifically, these interactions involve direct, face-to-face engagement, where robots are designed to detect human behavioral signals, with facial expressions being the most frequently analyzed signal [17, 18]. Additionally, much of the research in this area has been conducted using datasets that are publicly available or within the

controlled settings of a laboratory, a practice that is especially common when the studies involve physiological data for ER.

However, human-robot social interaction represents just one area of HRI applications. According to Thomas B. Sheridan [19], there are four primary applications within HRI including human supervisory control of robots, human remote control of robots, autonomous vehicles, and human social interaction with robots. To our best knowledge, there is no existing work that studies ER in the robot teleoperated scenario, which is one of the most important parts of HRI applications. The importance of ER in the human remote control robot HRI scenario cannot be overstated. However, the existing emotion sensing methods, including sensing human physiological and behavioral signals, show limitations for robot teleoperation scenarios. The common physiological sensing signals, such as heart signals and brain, are electrical signals which are sensitive to human motions, so acquiring these signals from operators in our system may add extra noise. Additionally, operators' behavioral signals, such as facial expressions and voice, are hard to capture in our system. More importantly, directly analysing human emotional signals faces similar limitations to those encountered in using human data for analyzing biometric identities, as previously discussed. It can only identify the emotional states at the access of the system. To address these challenges, we used robotic behaviors to classify operators' emotions enhancing the whole system's security. We developed one Dynamic Time Warping (DTW) based method and one Convolutional Neutral Network (CNN) based method to perform user-dependent and user-independent emotion classification respectively. This innovative approach not only circumvents the limitations associated with traditional emotion sensing methods but also contributes to the broader discourse on securing HRI applications through the nuanced understanding of operator emotions. The research path of this work is visualised in Figure 1.1.

1.2 Research Objectives

The main objective of this thesis is to secure teleoperated system by classifying users and user's emotion from motion-controlled robotic behaviors. The following research plan and tasks are considered to achieve three main objectives.

- Show that motion-controlled robots can inherit user biometric identities and users can be classified through the robots' behaviors. Build a foundation for the emotion classification work.
- Show that motion-controlled robots can inherit human emotions and these emotions can be classified by using robotic behaviors.

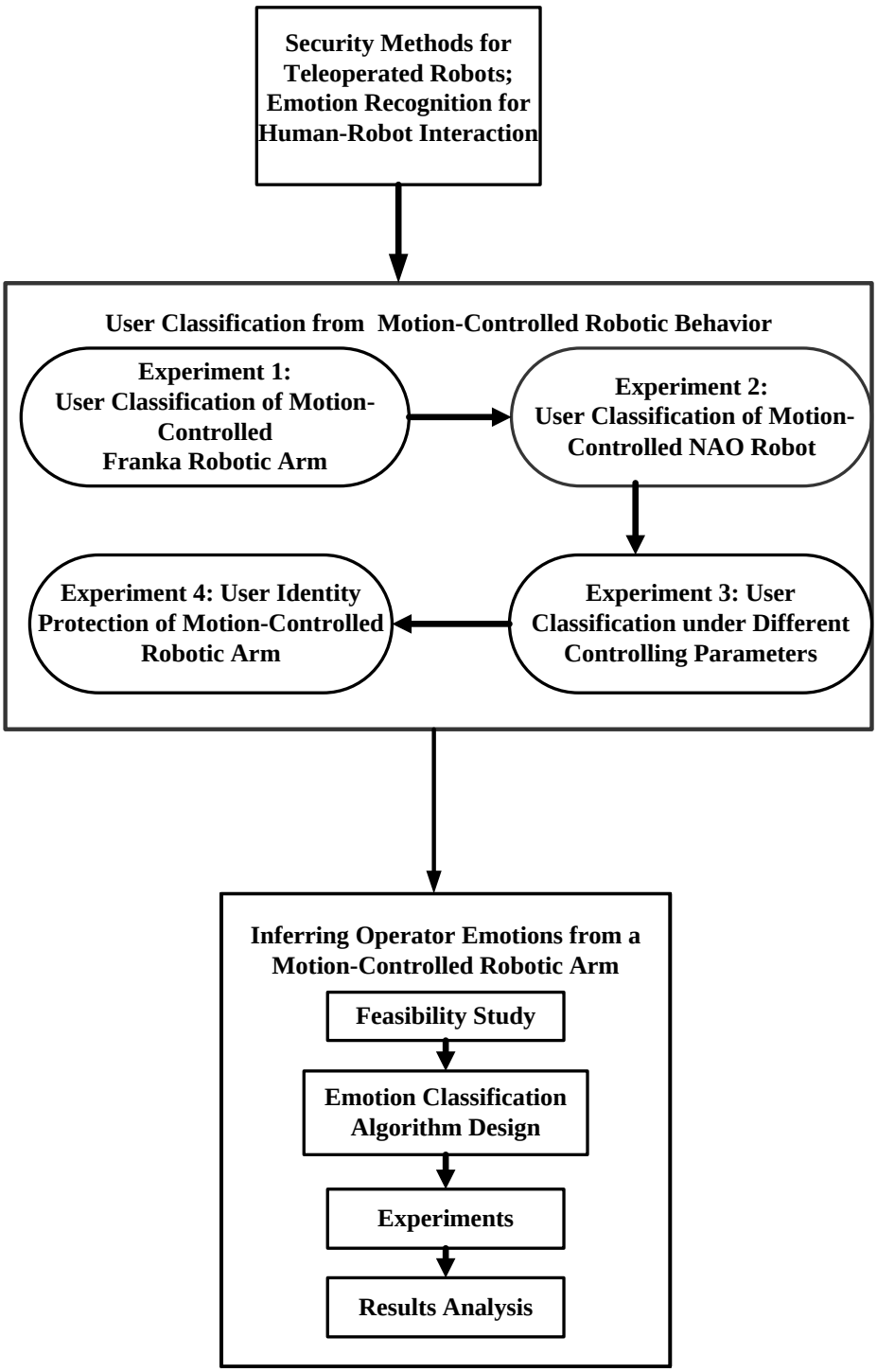


Figure 1.1: This flowchart depicts the research trajectory in this thesis. Each box represents different experiments and experiments are categorised into two topics. The connector means the order and connection between different experiments.

1.3 Research Contributions

Reflecting on the stated research objectives, this thesis has led to the following three research contributions:

- We verified that the motion-controlled robotic arm can inherit the operators' behaviors. Then, we implemented DTW and CNN algorithms on the robotic arm's end-effector trajectory data and showed that the users' identities can be classified. The ten user classification accuracy reached 95.0%. This methodology's adaptability was further evidenced through its application in real-world social tasks performed by a Kinect motion-controlled social robot, where user classification accuracy with NAO's two hands (accuracy: 93%) achieved results comparable to those with the Franka robot. Additionally, we explored the impact of varying three control parameters on user classification accuracy, discovering that parameters reducing control performance also diminished classification accuracy. This indicates that unsuitable control settings may result in the loss of valuable information, highlighting the potential for tailoring control systems based on individual user preferences. Finally, we proposed a reinforcement learning algorithm on the captured operator's trajectory data, ensuring the protection of user biometric information within the robotic arm's trajectory data. These studies established a solid groundwork for future ER research, offering a methodological and experimental design basis for forthcoming studies.

- We presented a comprehensive methodical review of ER in the context of HRI, with two primary aims. Firstly, we synthesized the current methodologies and applications of ER in HRI, providing a clear picture of the field's present state. Secondly, we identified promising ER techniques with potential future applications in HRI that have not yet been fully exploited. We contributed a unique taxonomy to categorise existing literature, distinguishing ER approaches based on various emotional signals, including physiological and bodily expressions. Furthermore, we positioned our examination of ER application in four pivotal HRI scenarios identified by the field: supervisory control of robots, teleoperated control of robots, automated vehicles, and social interactions with robots. We demonstrated the overall state-of-the-art techniques in most HRI related fields. For the supervisory and teleoperated control scenarios that have not been explored by ER, so we provided comprehensive information and practical recommendations for future designers. Our survey not only charted the current landscape of ER in HRI but also highlighted critical areas for future research, paving the way for more emotionally intelligent robotic systems.

- We demonstrated that the functional movements of a remote-controlled robotic avatar, which was not designed for emotional expression, can be used to infer the emotional

state of the human operator via a machine-learning system. We showed for the first time that user emotions can be accurately inferred from the movements of a motion-controlled robotic avatar. Then, We implemented two ER algorithms, based on DTW and CNN respectively, and developed unique emotional features from the avatar's end-effector motions and its joints' spatial and temporal features. Specifically, our system achieved 83.3% accuracy in recognizing the user's emotional state expressed by robot movements, as a result of their hand motions. Additionally, we demonstrated through direct comparison that our approach is more appropriate for motion-based telerobotic applications than traditional ECG-based methods. Furthermore, we discussed the implications of this system on prominent current and future remote robot operations and affective robotic contexts.

1.4 Thesis Organisation

In this section, we outline the organization of the remaining chapters of the thesis, structured as follows:

- Chapter 2, *Background and Literature Review*, research areas that inform this work and our objectives were reviewed. Firstly, we provided the background of teleoperated robots, existing controller types and robot types in the remote control scenario. Then, we provided the existing methods and their limitations for user classification. Based on this, it illustrated that user classification is one of the methods for robotic security. Next, it introduced the field of ER. It provided a brief introduction to emotion modelling that could be used for our work and briefly provided ER to the HCI and HRI fields. Then, it categorised the existing emotion sensing signals roughly into two types including the physiological signals and behavioral signals and shows the limitations of these signals used in our work. Furthermore, it provided existing methods of ER using signals from computer and robot interaction interfaces. Besides, it provided the existing approaches to endow robots with emotions and shows that robots have the ability to express emotions.

- Chapter 3, *User Classification from Motion-Controlled Robotic Behaviors*, described four experiments. The first experiment showed that a motion-controlled robotic arm can inherit human operators' biometric identities and user classification accuracy achieved to 95% using robotic end-effector trajectory data. Experiment 2 verified the versatility of experiment 1 proposed method on a motion-controlled social robot platform and user classification accuracy reached 93%. Experiment 3 showed that lower controlling performance leads to lower user classification accuracy. Experiment 4 proposed an AI agent to protect users' biometric identities.

- Chapter 2.2, *Emotion Recognition in Human-Robot Interaction: Taxonomy, Review, Current Challenges and Future Trends*, conducted an extensive review of emotion sensing and

recognition techniques, alongside a summary of current ER approaches in HRI applications. Building on this foundation, we outlined the prospective directions for ER within HRI. We also discussed practical solutions and challenges that could shape the future of ER in HRI settings. Based on this comprehensive review, we proposed our novel method of ER for motion-controlled robotic arm.

- Chapter 4, *Inferring Operator Emotions from a Motion-Controlled Robotic Arm*, for the first time showed that human emotions can be inherited from motion-controlled robotic behaviors and the ER accuracy using the robotic end-effector trajectory data reached 83.3%. We extracted 20 emotion-related features in total including the kinematic and expressive features. The ways to express emotions are different for different individuals, so we proposed a user-dependent method to classify emotions based on identified uses. However, classifying emotions without identifying users is more practicable in real-world scenarios, so we also proposed a user-independent method using a novel CNN method and the leave-one-subject-out-cross-validation (LOSOCV) accuracy reached 74.2%.

- In Chapter 5, *Conclusions and Future Works*, we discussed the research findings and limitations, summarised its contributions, and provided recommendations for future research along with the conclusions drawn from this thesis.

Chapter 2

Background and Literature Review

This chapter provides background information and presents a review of research in areas related to the research objectives in the thesis. Section 2.1 gives an introduction to the teleoperated robots. In addition, it introduces robot security methods and discusses the benefits and disadvantages applied to teleoperated robot systems. Section 2.2 gives an introduction to the emotion model gives a comprehensive review of emotion sensing signals and discusses future trends of emotion recognition methods for HRI.

2.1 Security for Teleoperated Robots

2.1.1 Teleoperation for Robot

HRI can be divided roughly into four areas of application, and teleoperation of robots is one of the most important parts of them. During the COVID-19 pandemic, as people were required to stay at home, teleoperated robots, which can bridge the physical gap between humans and remote environments, gained significant attention. A teleoperated robot is controlled remotely by a human operator [20]. Human provide control, while the robot is a follower followed by human control. It can be applied in the field of health [21], nuclear cleaning [22] and education [3].

Teleoperation Methods

Teleoperation methods for mobile robots can be mainly categorized into three types based on the controlling mode, including direct, supervisory, and multimodal teleoperation [1]. The first type is direct teleoperation, in which users provide direct control using traditional controllers, such as joysticks. The second type is supervisory teleoperation, where the operator provides high-level supervision. The third one is multimodal teleoperation, which is with an interface including multiple sensors, such as motion sensor, brain signal sensor and sound sensor. Among them, a wireless motion capture controlling interface that senses the operator's motions provides

mobility, safety and scalability. Among these methods, robot motion control using a motion capture device [3] provides flexibility and mobility for the users. In addition, it provides interactive and immersive experiences, enhancing user engagement and immersion. In addition to different teleoperation methods, there are several types of robots that can be used for teleoperation in different scenarios, including robot hands (e.g. Shadow Robot Hand), robot arms (e.g. Franka Robot Arm), social robots (e.g. Nao), and humanoid robots (e.g. Boston Dynamic Dog).

2.1.2 Security Methods for Teleoperated Robots

A teleoperated robot system contain muiltiple devices, such as sensors, controllers and robotic devices, and refers to long distance data transmission, during which the system is susceptible to both cyber security threats and physical threats. Securing the teleoeprated robot system can enhance the human-robot trust, which is crucial in today's world where modern social robots are increasingly being deployed [23]. When the robot system is under attack, the robot is not safe and the system is highly possible to produce negative collaboration outcomes in remote environments.

Encryption

Teleoperation requires communication of data between the operator and the remote robot, which may be subject to cyber attacks [24]. Encryption can prevent these attacks by converting regular data to incomprehensible cypher text [25]. However, these security measures require large computing resources and create additional overheads in communications, which adds latency to the operation and impacts the operator's ability to effectively operate the remote robot in real time [25].

Robot Misbehavior Detection

In order to address the security issues on robot devices, robot misbehaviors detection techniques are employed to detect the abnormal and unusual behaviors of robotic devices. The common used method is learning-based machine learning and deep learning algorithms to detect unusual signals [26, 27]. However, this method is used for autonomous robotic systems and cannot protect the robot when there is an impersonation attacker and cannot detect spoofing attacks.

User Authentication

User authentication is a method that confirms the identity of a user while accessing a computing device, such as a mobile phone and laptop and teleoperated robot systems, or an online service,

such as an email [28]. A wide variety of authentication methods have been introduced, which can generally be categorized into three main types including knowledge-based, token-based and biometrics-based [29]. The knowledge-based method is to store passwords in advance, such as text passwords, and then the user verifies the ownership of the passwords to log in. It is easy to use, however, it is susceptible to theft and anyone who knows the text passwords can get authenticated. Besides, token-based allows users a pre-aligned tokens, such as a mobile phone, a key and a smart card. They do not need the user to remember, however, the methods are burdensome and also susceptible to theft. Lost or stolen tokens can easily enable unauthorized individuals to pass authentication [29]. Finally, for biometric-based method, relies on inherent biometrics factors of human users. This method requires the user to provide multiple samples of either a physical (e.g., fingerprint) or a behavioral (e.g., gait pattern) trait. The techniques are to classify different users using extracted features. However, traditional physical biometrics (i.e., fingerprint and face recognition) are vulnerable to spoofing attacks [28]. For example, fingerprints can be easily reproduced from a photograph [30]. Similarly, facial recognition can be deceived by using a photograph of the victim found on social media. Behavioral biometrics is a relatively new concept using the data acquired from sensors on personal smart devices [30]. For example, an accelerometer and gyroscope on a smartwatch can record the specific arm gesture data that are used for user authentication [31]. It can allow for continuous authentication, which lasts for a long time after user-login. However, the existing behavioral biometrics authentication can only protect the security of the access point of the teleoperated robot system and can not guarantee the whole system's security during the data long-distance transmission. In addition to use single modality to authenticate users, multi-Factor authentication (MFA) is to use more than one authentication mechanism to provide a high level of safety [32]. For the most part, MFA is based on users' physical and behavioral biometrics information [32]. For example, user wear a ring for token-based authentication, to perform gestures for behavioral biometraic authentication. However, existing MFA methods can only protect the teleoperated robot system partially (user access point), which can not guarantee data are not tampered with during the data long-distance transmitting.

2.2 Emotion Recognition in Human-Robot Interaction: Taxonomy, Review, Current Challenges and Future Trends

2.2.1 Introduction

Emotions play an essential role in human communication and social interaction, influencing and shaping human behaviors [33]. Human emotions can be recognized using a variety of information sources, ranging from exterior physical signals and movements to internal physiological signals. With advancements in emotion recognition (ER) techniques, automated emotion recog-

nition (AER) has begun being applied in different areas, such as health [34], education [35], security [36], marketing [37], entertainment [38] and robotics [17]. ER has also seen increased development in fields such as human-computer interaction (HCI) [39], virtual reality (VR) [40], augmented reality (AR) [41], advanced driver assistant systems (ADASs) [42] and human-robot interaction (HRI) [43]. The most prominent sensors and techniques used in emotion detection include electroencephalograms (EEG), electrocardiograms (ECG), electromyography (EMG), cameras, and recorders. Building emotion recognition systems for HRI can further involve the adaption of techniques including Computer Vision (CV), Machine Learning (ML), Deep Learning (DL) and Natural Language Processing (NLP).

In order to enhance the interaction and achieve symbiosis between human and robot [44], it is critical to facilitate emotionally intelligent HRI systems. Emotional intelligence is essential to recognize, infer, generate, and express the emotions [45], [46], [43], leading to improved communication [47], personalized interactions [48], enhanced user experiences [48], adaptability [1], and ethical considerations [49]. By incorporating emotion recognition, robots can foster more engaging, empathetic, and harmonious relationships with humans, which is important for robots performing complex tasks in social environments which require close interaction and cooperation with humans. For example, social robots in hotels provide guests personalized services with help of emotion recognition [48]. In addition, inferring passengers' emotions can improve user experience and safety in AVs [2]. Similarly, industrial robots can dynamically shoulder more workload when detecting workers' fatigue from the view of ethical considerations [49]. In terms of mechanisms, robots can utilise human behaviors, such as touch [50], bodily movements [51], facial expressions [52], physiological signals [53] and signals from interaction interfaces [54] to recognize human emotions for enhanced interactions by responding with gestures, facial expressions and voices [55] or giving warnings and adjusting paths [56].

Related Works

In the existing literature, there are three main categorisations of ER. The first category of works reviewed the single-modality or multi-modality emotion sensing methods, such as ER using bodily movements [57], facial expressions [58], and combination of visual and audio signals [59]. The second category reviewed and summarized sensing and recognition algorithms for ER. For example, Dzedzickis *et al.* [60] summarized different sensors and corresponding algorithms of ER under these sensing methods, and Wang *et al.* [61] reviewed the emotion dataset and surveyed detailed architectures and performances of unimodal and multimodal ER. Finally, the third category of works surveyed the existing methods of ER for interaction applications. For example, Cowie *et al.* [39] reviewed the ER that used speech signals and its application application to entertainment robots. Although ER for HRI is a key capability for potential future HRI

[1] https://www.suaave.eu/

[2] https://cordis.europa.eu/project/id/815003

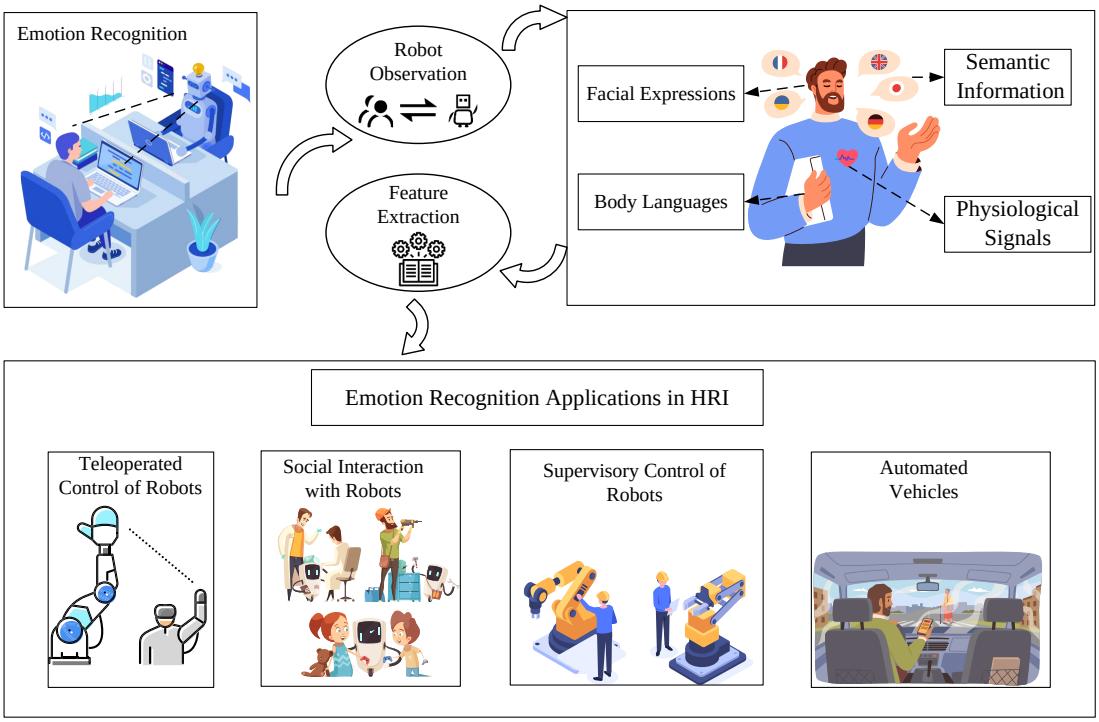


Figure 2.1: Emotion recognition in human-robot interaction.

applications, it is relatively unexplored topic with only one survey study exploring ER applications in the field of HRI [43]. In the study, the authors summarized the existing ER methods for HRI applications, but only in the context of human-robot social interaction. However, there are four main categories of HRI application, including human supervisory control of robots, human remote control of robots and automated vehicles, alongside human–robot social interaction with robots [19], which require further exploration.

To address this gap and answer the essential question of how future HRI designers can utilise ER in emerging applications, we present a comprehensive review of proposed emotion-sensing methods with more specific branches. In addition, we review existing ER methods in HRI applications and discuss future trends and challenges, providing insights and recommendations on how and why ER could be applied to benefit a variety of HRI scenarios. For each HRI applications, we individually review the existing ER methods and discuss the potential challenges and practicality of using other ER methods. Furthermore, we provide potential ER methods that can be applied to specific HRI applications in the future. The overview is shown as figure 2.1.

2.2.2 Background

The importance of emotion recognition in HRI applications

Sheridan *et al.* identified four main applications of HRI [19], categorized through human involvement factors and robot automation during the interaction. The first category is Remote

Robot Control which requires human intervention at all times, without any means of autonomy. The second is the Human Supervisory Control, which allows the robot to do some routine tasks, such as assembling, delivery, picking, and placing. During the processing, the robot is half programmed and half autonomous, endowed with intelligence, such as predicting the operator's trajectory and avoiding the collision. The third category is Automated Vehicles, where the robot is entirely autonomous. The last category is Social Robots which are automated but also express emotional intelligence, such as emotions and empathy, to engage in social interactions with humans. ER can offer essential or beneficial utility to each of these HRI application areas. Psychologists have shown that emotions are a fundamental aspect of human intelligence and rationale, and that intelligent behaviors are influenced by emotions [62]. Emotions constitute the primary motivational system for human beings [63], thus by identifying different emotions, a person's intrinsic motivations can potentially be inferred. In [64], authors showed that frustration is an indicator of human interest during the interaction with a robot and can help to understand whether the interaction is successful or not. In addition, emotions can be an indicator for mental health concerns and ER can be used in the diagnosis of conditions such as depression [65] and chronic stress [66] based on detection of negative valence and positive arousal states [67].

Emotion recognition is an interdisciplinary field involving psychology, computer science, and signal processing, with many application areas such as robotics [55], health [68], security [69] [70], education [71], website customization [72] and marketing [73]. ER can benefit HRI in five core ways.

1. First, ER allows robots to provide personalized interaction and services with humans. For example, researchers in [48] evaluated service robots in hotels and showed that although robots showed superiority in repetitive tasks, they cannot provide personalized services, which requires social ability and emotional intelligence.

2. Second, emotions play an essential role in empathetic human communication and social interaction. ER can lead to more natural and effective human-robot communication by enabling robots to perceive human emotional states and respond accordingly, for example by offering adjusted support, comforting actions, or giving the user some space [44].

3. Third, ER can improve the robot's adaptability during scenarios where strong emotions can have a negative impact. For instance, when a telerobot detects intense or suppressed emotions in surgery, the robot could warn the operator to avoid imprecise or exaggerated operations that may cause severe injuries to the patient.

4. Fourth, ER could be used to automatically enhance and optimize user experience and satisfaction. For example, in automated vehicles, the system could adjust the temperature and music inside the cabin according to the inferred emotion. In addition, the trust and

adoption of passengers could be evaluated based on inferred emotion, providing implicit feedback to the system.

5. Fifth, ER can facilitate ethical and responsible interactions in settings with co-present robots and humans. For example, in human supervisory control of a robot the operator and robot cooperate in the same workspace to achieve industrial tasks, such as assembling and delivery. Recognizing human emotions could facilitate a more human-centered and trustworthy work environment as the robot takes on more workload, initiates stoppages or calls for support in response to human distress as a result of overwork or a workplace policy violation.

The Definition of Emotion Modelling and Methods of Elicitation

Psychologists Scherer et al. [74] illustrated that emotion models can be categorized into three types: discrete emotion, dimensional, and componential. The discrete model defines emotions according to distinct class labels, which can be six basic emotions: anger, disgust, fear, happiness, sadness, and surprise, as introduced in [75]. The most prominent dimensional model is Russel's Circumplex Model of Affect [76], featuring two dimensions: valence and arousal. Valence represents how positive or negative the emotion is, while arousal refers to how high or low the physiological arousal occurring during the emotion. Plutchik et al. [77] introduced the best-known componential emotion model, a hybrid of the first two models, where emotions are represented hierarchically.

Eliciting emotion is a foundational requirement for ER research and is inherently challenging as the participant's observable emotional state is ambiguous and may change during data collection. Various methodologies have been explored to elicit specific emotions from participants. Some studies employed a subject-elicited approach [62], such as using actors to express emotions [78]. Others employed an event-elicited approach [62], seeking to elicit emotion from participants through evocative songs, video clips, text and pictures, using resources like the International Affective Picture System (IAPS) [79]. However, gathering high-quality data has strict experimental environment requirements, thus the majority of work has utilised pre-existing open-source high-quality datasets. One of the most popular physiological datasets is DEAP [80] which includes EEG and peripheral physiological signals, but there are many other datasets varied by particular collection methodology or control variables. Katsigiannis and Ramzan [81] contributed a dataset of ECG and EEG signals in response to emotions elicited using audio and visual stimuli collected in an isolated environment. Datasets also vary by participant movement, Koelstra et al. [80] elicited emotion from seated participants, while Busso et al. [82] using moving participants. Likforman et al. [83] conducted a handwriting dataset, while Gunes et al. [84] contributed a face and body dataset. However, datasets are typically collected in laboratory conditions, far removed from real-world environments [60, 61]. Consequently, there is a growing need for ER experiments conducted in real-world scenarios, especially in the field of HRI.

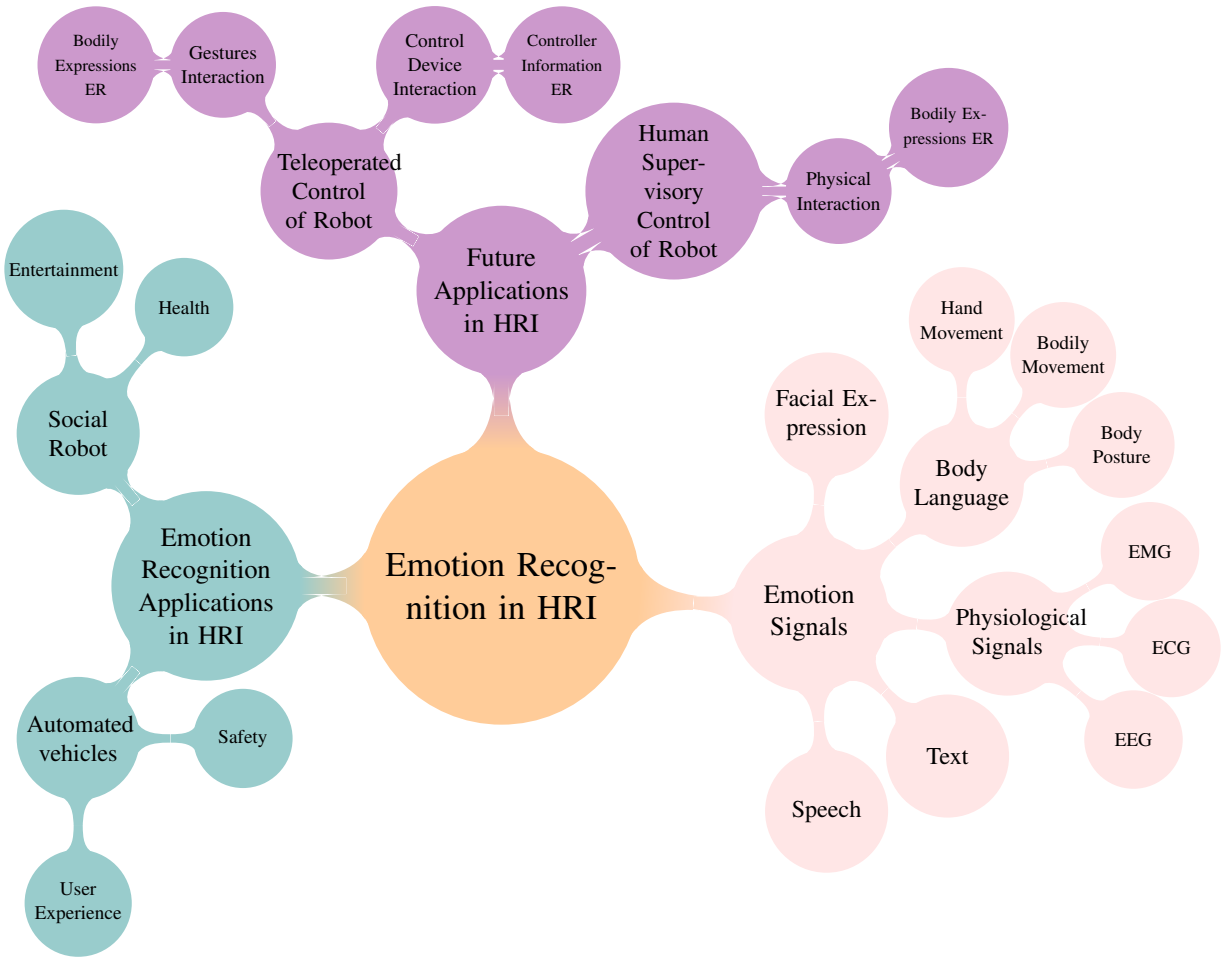


Figure 2.2: Taxonomy structure of emotion recognition in human-robot interaction.

2.2.3 Taxonomy

In this section, we provide a taxonomy of emotion recognition in HRI, as shown in Figure 2.2. ER in HRI is an interdisciplinary field and includes robotics, artificial intelligence, and psychology. In this survey, we review the existing work of emotion recognition in HRI applications and propose its potential development in future. The HRI applications are mainly categorized into four areas [19], including remote control, human supervisory control, automated vehicles, and human-robot social interaction. We survey the emotion recognition work for the HRI scenarios that have been explored, including human-robot social interaction and automated vehicles. For the HRI application that has yet to be explored in ER, we analyze whether these areas would benefit from ER, whether the existing ER method can be applied in the future, and how to apply them.

Emotion Signals

The existing studies of emotion signals are categorized into physiological and physical data. Physical data includes facial expressions [58], speech [85], bodily expressions [57] and text [86].

Physiological data includes brain activity [87] and peripheral physiological signals, such as Electrocardiography (ECG), Heart Rate Variability (HRV), Skin Temperature (SKT), Respiration Rate (RR), blood volume pulse (BVP), Skin conductance (SKC), and Galvanic Skin Response (GSR) [88]. In addition, these unimodal data streams can be combined to allow for holistic multimodal affective signal analysis [61, 89].

Existing Emotion Recognition in HRI Applications

Existing studies of ER in HRI have mainly discussed the changes in a human's emotion when they socially interact with robots [90–94] to help design interaction interfaces, enhance interaction, monitor excise, assist treatment of children impairment and relieve loneliness of the elder. Additionally, there exists work of ER for automated vehicles [47, 54, 95–100], where it can promote the passengers' mental health, understand the feelings of passengers' acceptance, satisfaction and trust. ER sensing methods for AVs include facial expressions, voice, physiological signals and motions of passengers.

Future Emotion Recognition in HRI Applications

ER is necessary to facilitate more interactive and intelligent systems in real-world scenarios for HRI applications. In human remote control of robots in hazard environments, ER can improve system security. For example, the detection of intense emotions could avoid imprecise operation. In human supervisory control of robot in industry, ER can help to build trust between human and robot, as well as evaluate workers' satisfaction. While the use of ER in these areas is unexplored, there are several viable emotion sensing methods which could be implemented. Physiological signals can be used, but it can be challenging to acquire accurate data in human remote and supervisory control of robot scenarios, due to data artifacts caused by human movements [101]. Besides, acquiring physiological data can require on-the-body sensors or wearables, which could present a practical challenge. Compared with the methods of acquiring physiological signals, observing physical signals is more practical, as physical data can be acquired from different task scenarios. For example, ER can leverage facial expressions or the operators voice, while in scenarios where operators do not talk but make hand moves, these hand movements can be analysed to recognize emotions [102]. Another example is using information from interaction interfaces, such as haptic controllers [103].

2.2.4 Emotion Signals

Human emotion states can be detected from changes in both internal physiological signals and outer behaviors' changes. Inner information is instinctive and often less overt, while behavioral information, including vocal, facial, and body information, are primary methods of human interpersonal communication [58]. The following section will discuss existing ER sensing methods,

Table 2.1: ER methods categorization

Methods		**References**
Peripheral physiological signals	Contact sensors	[60, 61, 68, 104–109]
	Non-contact sensors	[78, 110, 111]
Brain signals		[43, 112, 113, 113–116]
Facial expressions	Visual Information	[58, 117–122]
	Thermal information	[100, 123, 124]
	EMG signals	[125]
	Eye tracking	[126, 127, 127, 128]
Speech, text and handwriting	Speech	[85, 129–131]
	Text	[86, 86, 132–134]
	Handwriting	[135, 136]
Bodily expressions		[57, 60, 102, 102, 137–145]
MultiModality	Facial& Body	[120]
	Audio & Visual	[146]
	EEG & EMG	[147]
	Facial & Body & Audio	[148]
	Text & Speech & Visual	[149]
	Facial & Hand & Body	[150]

as summarised in Figure 2.1, in more detail.

Peripheral physiological signals

The autonomic nervous system (ANS) is a general-purpose physiological system [151] and psychologists [152] [152] have shown that emotions influence the activity of the ANS. Physiological responses can be measured from peripheral physiological signals to facilitate emotion recognition. Relevant peripheral physiological signals include Electrocardiography (ECG), Heart Rate Variability (HRV), Skin Temperature (SKT), Respiration Rate Analysis (RR), Blood Volume Pulse (BVP), Skin conductance (SKC) and Galvanic Skin Response (GSR) [88]. Methodology for measuring these signals can be broken down two main categories: sensors that require contact and contact-less sensors.

- **Contact sensing signals:**

  ECG is most prevalent among the peripheral physiological emotion signals used for ER. The autonomic nervous system (ANS) connects the brain and heart [153] and brain behaviors influences heart behaviors, allowing the detection of emotions from heart activity [104], [68], [154], [105], [106]. The primary methods of ER using ECG utilise Machine Learning (ML) and Deep Learning (DL) [61]. For ML-based ECG emotion recognition, features are extracted manually. The detected ECG signal is a wave from which five crucial components that contain information about heart activity can be extracted: the P, Q,

R, S, and T values [105]. Time-domain, frequency-domain, and nonlinear-domain analyses of these features are conducted. ML-based classifiers are used on these features for emotion recognition, with the SVM classifier being prevalent [61]. For DL-based methods, [106] self-supervised learning is used to detect ECG transformations and transfer the trained parameters into the classification network. ECG signals are, however, sensitive to human movements and therefore can be challenging to collect outside of a controlled environment. Furthermore, this method requires a large amount of data and can not be used for instantaneous ER [60], all of which impacts its usefulness in real-world applications. Another signal which can be used for ER is Skin Temperature Measurements (SKT), using biomedical sensors attached to human skin to detect temperature changes [108].Emotions, such as like stress, anxiety and anger, can cause a decrease in finger temperature. The utility of SKT is, however, impeded by a long latency between temperature change and emotional stimulation, additionally, it cannot be used to recognize precise emotions [60]. Respiration rate (RR) can also be used for ER [109]. It suffers from limitations, as RR is influenced such as the human movement and surrounding environment which could confound ER [60]. In addition, unlike SKT, RR can also be intentionally changed by the user, which could result in less accurate data. Finally, the quality of the data acquired by using contact sensors depends on the quality of the sensor and suffers from the movements of the users, thus all these methods may be impractical depending on their use in different real-world HRI applications.

- **Wireless sensing signals:**

  In addition to using contact sensors, prior work [78, 110] has utilised wireless devices to detect and extract heart activity and achieved comparable ER results using an ECG sensor. For example, SKT signals can be observed using infrared cameras to capture thermal imaging of skin temperature [111, 155]. While contact sensors can be cumbersome, wireless signals are more convenient and do not influence or impede normal human activities. However, current research has utilised strict experimental environments, thus the ability for wireless methods to achieve precise signal extraction despite naturalistic human interaction or movements has not been demonstrated. The wireless sensing methods do not require physical contact, which is beneficial in environments where contact is challenging or sensitive and provides flexible deployment. However, due to Wireless signals can be affected by physical obstructions, such as wall, and environmental interference, such as electromagnetic disturbances. Besides, some of wireless sensors have lower accuracy and resolution compared to contact sensors [156].

Brain signals

Prior work [112] has shown that by observing brain activity, it is possible to correlate physiological responses and emotions even more effectively than by observing peripheral physiological activities. Electroencephalography (EEG) is a widely used non-intrusive technique to detect brain electrical activity. The features are extracted from time-domain, frequency-domain, time-frequency domain using electrodes [113] [114] [115]. The predominant features of EEG are frequency features from different frequency bands, including delta, theta, alpha, beta and gamma, in which the power of each band, the statistical features of different power spectrum, and Higher Order Spectra (HOS) features have been explored [60, 113, 157]. Suhaimi et al. [116] summarized the popular machine learning algorithms for ER using EEG in recent years: K-nearest neighbor (KNN), support vector machine (SVM), and artificial neural network (ANN). Others have utilised a deep learning approach, with the best recognition accuracy (90.40%) achieved by Zhong et al. using a Dynamical graph convolutional neural network [158]. EEG has two advantages over other methods. First, it can reach a higher emotion recognition performance, and second, it is feasible to use a commercial-grade EEG device for ER [43], rather than specialist equipment. However, most work has focused on developing the method and using the published EEG dataset DEAP. Furthermore, as EEG is an electrical signal is it sensitive to human motions and influenced by motion artifacts [159], which could limit its applicability in non-controlled interactions with robots, as with previously discussed methods.

Facial Expressions

The increase in readily available graphical computational processing power and variants of neural networks has lead to increased viability of facial emotion recognition (FER) approaches [160]. There are two approaches for FER: traditional FER and neural-networks-based FER. Traditional FER utilises the general computational flow with five steps: data input, signal preprocessing, feature selection, and classification [160]. In [161], researchers provided a detailed data processing method and network design. FER has been applied in real-world applications. For example, emotions observed using FER are used to evaluate the user satisfaction of tourists in [162].

- **Visual information:**

  Facial expressions are direct and natural signals, and primary information channels in human interpersonal communications [58, 118]. With the development of the computer vision field, more FER methods have been explored in recent years. Several emotion facial datasets have been contributed and researchers have proposed deep learning algorithms using these datasets [121, 122]. There are two ways of using visual information to recognise emotion. One is to extract features manually, and the other is that deep neural network output features automatically [58]. Features are calculated from action units

(AU) and landmarks. AU was proposed in [117] and represents muscle movement. Landmarks represent the facial characteristic points, such as the nose and eyebrow positions. Tarnowski et al. [118] used a Kinect to capture face characteristic points and build a 3D model. Then, AUs are calculated from the selected facial points, and K-Nearest Neighbors (KNN) and multilayer perceptron (MLP) are used to classify seven emotions. Authors in [119] used 51 facial landmarks and then extracted the features and used SVM for classification, deployed on a mobile application. In [120], researchers extracted features by handcraft and proposed temporal segment detection methods to process video information. For automatic selection, Jain et al. [121] proposed a CNN and RNN combination method on the JAFFE-faces dataset. In [122], they used VGGNet architecture on the FER2013 dataset and achieved to highest accuracy of 73.28%. However, there are a few drawbacks of using facial visual information. Firstly, based on deep learning characteristics that require a large amount of data, FER needs abundant computing resources. In addition, the range of human movement can be limited when collecting facial data. Additionally, collecting facial data involves serious privacy issues. More importantly, micro-expressions are difficult to identify due to their spontaneous and subtle nature of occurring involuntarily, and new evaluation metrics are needed to observe micro-expressions from moving images [58].

- **Thermal information:**

  The visual information of facial expressions is detected over the skin, so it is vulnerable to environmental factors, such as occlusion of external sensors. The thermal information of the face can be measured instead [163] via on-the-skin sensors, which can circumvent such confounding factors. In [100], they used handcraft manual feature selection, while Shaees et al. [123] used transfer learning to perform FER on a thermal facial dataset. Additionally, Nayak et al. designed a two-stage approach to recognize emotions through time-series thermal sequences including Regions of Interest (ROI) detection using RNN and emotion recognition using DTW [124]. The fusion of facial visual and thermal information has been applied in face recognition [164, 165], which can be further explored to facilitate FER. Environmental temperature can, however, play a confounding role when sensing thermal information, a key limitation [43] for its use in ER. Related experiments [166] were conducted in climate-controlled laboratory settings, while temperature is not controlled in different daily life scenarios. Besides, small sized thermal imaging systems are needed to provide mobility [166].

- **EMG signals**

  Rather than using visual information, an Electromyogram (EMG) can be used to detect facial muscle signals. In [125], authors adopted wavelet packet transform on facial EMG signals and used SVM to classify the extracted features. However, EMG sensing methods

are in the experimental stage [167, 168] and there is not any application to ER, making this a possible but unexplored avenue.

- **Eye tracking (ET)**

  Eye tracking devices can detect a person's eye movements and position [169]. There are different implementations, including desktop eye-tracking (Gazepoint GP3 eye-tracker), mobile Eye-Tracking (Tobii Pro Glasses 2 eye-tracker) and Eye-Tracking in Virtual Reality head-mounted displays (HMD) [128]. Eye behaviors can reveal crucial information related to emotions and cognition processing by observing features such as eye movement, gaze patterns, motion speed, pupillary responses and fixation duration [127, 128]. In [127], they extracted 18 features to classify 3 emotions with an accuracy of 80%. In most cases, eye tracking is combined with EEG to classify emotions [170, 171]. Plopski et al. [126] highlighted how ER facilitated by ET in Extended Reality can lead to nuanced and richer gaze expressions in both remote and co-located collaboration [126]. There are some downsides to this approach, however, as the features of pupil diameters are influenced by lighting conditions, ET is mostly only used in multimodal ER and it requires either worn hardware or fixtures which require specific positioning of users.

Semantic Information

Semantic information refers to the meaning and interpretation of words, phrases, and sentences and it is a key concept in ER, mainly including speech, text and handwriting.

- **Speech**

  Speech is another natural method of interpersonal communication and speech emotion recognition (SER) has been developed over the past decades [85]. SER shows particular benefits in scenarios where there is no other signals to observe for ER, as it cannot be visually occluded or confounded by movement. There are two main approaches: traditional and deep learning techniques [129]. Traditional techniques include feature extraction, feature selection, and classification. There are two categories of features: short-period features such as formants and pitch, and long-period characteristics, such as intensity and variance. Then these extracted features are fed into a classifier, such as SVM and KNN. Results in [130] showed that deep learning techniques perform better than traditional techniques in SER. In [130], authors used the transformer (a machine learning model) to recognize emotions in a conversational context. Based on the characteristics of the transformer, it can learn the immediate previous information in a sequence, instead of learning current or nearest information, which yielded state-of-art results. Self-supervised learning can also be used in SER to learn emotional signal representations. Atmaja et al. [131] used self-supervised on different public speech dataset and showed the accurate relationship between emotions and classical acoustic features.

Speech signals are the main semantic information in the scenario of HRI, especially in the social robot interaction scenario. ER using speech signal can augment reliable, trustworthy and low-latency communication between robot and human. For example, a robot could provide emotional support corresponding to human emotions when talking with human, which is naturally interpersonal communication and mirrors the empathetic exchanges found in human-to-human interactions. Thus, it allows human to build more trusting relationships with robots.

However, using SER methods requires a conditional experiment set up and implementation in the robot can add noise from the robotic engine. Authors in [172] proposed a DNN method to improve the robustness of SER in a real robotic settingby using several data augmentation techniques so that the model can resist noise from robots, rooms and acoustic events.

- **Text**

  Emotion recognition from text is one of the most significant and challenging natural language processing (NLP) tasks [86]. There are five approaches to speech ER: keyword-based, rule-based, classical learning-based, deep learning-based, and hybrid [86]. The keyword-based method finds the occurrences of keywords in the text and assign emotion labels corresponding to keywords. The main steps of rule-based text ER are rules extraction and selection. For classical learning-based, researchers extracted the features through human speech and used these features to feed machine learning classifiers to recognize different emotions. In addition, deep learning is also used in this field. The most used deep learning model is long short-term memory (LSTM). The hybrid method combines the first four approaches, such as a combination of keyword-based and learning-based methods [132] and rule-based and learning-based [133]. According to Alswaidan et al., [86], keyword-based methods are mostly used in a text to recognize explicit emotions, while the others are for implicit emotions. Using a transformer can solve these problems because it learns information from previous content. For example, prior work [134], adopted bidirectional encoder representations from transformers (BERT) to recognize emotions considering contextual information. However, text ER is mainly used in human-computer interaction (HCI), where text data is directly acquired, while text is harder to acquire in HRI scenarios.

- **Handwriting**

  Handwriting is controlled in real-time by a person's brain and research has found that a person's emotions, mental health, and other personality traits can be conveyed from handwriting analysis [135, 173]. The primary method of handwriting emotion recognition (HER) is traditional machine learning. Features such as Slant, baseline, pen-pressure, size, margin, and zone [135] are extracted manually, as well as stroke-related features, such as

stroke duration and length [136]. Then, these features are fed into a classifier. HER can be used in mobile and ubiquitous computing or depressed detection [135, 136]. However, it is not a widely discussed topic and few works studied it, with the most accurate detection rate being 70%, achieved in [136]. However, it has similar real-world implementation problems with text methods, which is not directly acquired.

Bodily Expressions

Mauss et al., [112] introduced that physical human behaviors, including facial displays, vocal information, and bodily expressions, are primary ways to convey emotions between interpersonal communications. Although bodily expressions are a less discussed topic, Ahmend et al. [102] suggested that ER from analysis of bodily expressions has huge potential to revolutionize robotics by impacting human interaction behaviors. Bodily expression information can be observed from the head, hands, torso, upper body, or whole body. The general pipeline of ER for bodily expressions includes body detection, human body modeling, feature extraction, and emotion recognition. The measurement techniques are non-contact techniques, including a depth camera (e.g., Kinect), RGB camera, and motion capture system (e.g., Optitrack, Vicon [143]). Previous studies [57, 102, 137–139] have used bodily movements to recognize different emotions. Glowinski et al. [137] introduced the analysis of upper body features, including expressive and dynamic features. Expressive features include energy, spatial extent, smoothness, symmetry, and head tilt. Except for the expressive features, 25 dynamic features can be considered, including the variance, peak duration, main peak duration, and Number of Maxima from video data. In [137], the authors used the displacement of major joints, motion features, and the Laban Movement Analysis (LMA) effort component and mass displacement. In [140], CNN and RNN were used to analyse basic discrete emotions using body skeleton movements that are built using main body joints, while others have recognized continuous emotions based on the features from LMA [141]. Most prior studies have used stylized motion tasks (tasks that explicitly are designed to express different emotional states) [137] [102] [138], but some studies have also demonstrated the potential of using non-stylized motions, i.e., functional movements to accomplish tasks irrelevant to emotional expression [142–144]. In [144] and [143], authors analyzed non-stylized gait to recognize emotions using the captured skeleton movement. Additionally, according to Ding et al. [145], hand gestures can also be used to classify emotions, while Dzedzickis et al. [60] asserts that bodily expression is a promising approach in future practical ER cases, due to its wide applicability.

Multimodal information

As emotion information can be conveyed from many different sources, combining more than two sensors or modalities can allow for more information and a higher resulting recognition accuracy. Prior work has explored several signal combinations. The main combination types

are different physiological signals combination, physiological and behavioral signals combination, and behavioral signals combination. Prior work [120] has combined facial expressions and bodily displays and showed that the performance of feature-level fusion (early fusion) is better than the decision-level fusion (late fusion). Facial expressions, hand gestures, and body postures has also been combined together to classify students' emotions in classroom environment [150]. Others [148] have combined facial expression, body gestures, and acoustic information in a speech-based scenario where humans communicate with the agent. Previous research [174] designed deep neural networks to learn facial, body, and audio information. Self-supervised learning was used to select features from text, speech, and vision signals that were fed into the transformer [149]. Siriwardhana et al. [149] collected the features of text, audio and vision signals, then used a transformer to recognize different emotions, while Schoneveld et al. [146] leveraged a deep learning method on audio-visual emotion recognition. Zhu et al. [147] used EEG to evaluate customer preference and used eye tracking to fine-tune parameters at the application level. The efficacy of EEG and ECG signals when isolated or combined has also been explored, showing a 35.78% increase in performance when used in combination [175]. Despite the upsides of these multimodal approaches, there are some disadvantages and limitations. These approaches may require that more data is collected, raising greater concerns over privacy issues. Additionally, the cost and complexity of these approaches will be higher as a result of combining or coordinating multiple pieces of hardware and data-streams.

2.2.5 ER in HRI Applications

There are two types of HRI application where the use of ER has been explored: Automated Vehicles and Social HRI (see Table 2.3).

Emotion recognition for automated vehicles

In this use case, the robot is an automated vehicle (AV) that makes decisions and judgements by itself, while the human is primarily just a participant. ER in AVs is an emerging area of research that aims to understand how people perceive and respond to AVs from an emotional perspective, which improves user experience and supports the acceptance of AVs [3]. A comprehensive survey on driver emotion recognition [96] showed that the facilitation of ER in intelligent vehicles potentially promotes passengers' greater mental health and as AVs become more prevalent there is a growing need to use ER to understand passengers' feelings of acceptance or satisfaction inside the car [4]. For example, anxiety passenger anxiety could be used to indicate dissatisfaction with the experience. By recognizing passengers' emotions that indicate their feelings of willingness, trust, comfort, and safety, the AV can adapt and adjust its behaviour. The EU has funded

[3] https://cordis.europa.eu/project/id/815003

[4] https://projects.research-and-innovation.ec.europa.eu/en/projects/success-stories/all/building-automated-vehicles-are-tune-your-emotions

Table 2.2: State-of-the-art emotion classification methodologies encompass a comprehensive analysis of employed emotion sensing techniques, algorithms, method performance, and the types of emotions identified.

Emotion Signals	Methods	Performance	Emotions	Ref.
ECG	Self-Supervised Learning	Averaged accuracy: 96.15%	Arousal and valence	[106]
EEG	Dynamical graph convolutional neural network	Accuracy: 90.40%	Negative, neutral and positive	[158]
Facial Expressions	CNNs: VGGNet architecture	Accuracy: 73.28%	Anger, disgust, fear, happiness, sadness, surprise and neutrality	[122]
Eye Tracking	17 Features extraction SVM classifier	Accuracy: 80%	C1:high arousal and high valence C2: low arousal and moderate valence C3: high arousal and high valence	[127]
Semantic Information	Transformer	WAA: 68%	Happiness, sadness, neutral, anger, excitement and frustration	[130]
Body Language	Generalized Zero-Shot Learning (GZSL)	Accuracy: 72.92%	Happiness, sadness, surprise, fear and anger	[176]

emotion recognition in AVs to enhance user acceptance, such as Trustonomy [5], DriveToTheFuture [6] and SUaaVE [7]. By understanding the emotional responses of users to AVs, developers can design systems that are more appealing and user-friendly, or this could allow AVs to respond accordingly to passenger emotions.This can be important for providing a more personalized and responsive driving experience. If the autonomous vehicle detects that the passenger is feeling anxious or stressed, it has the ability to modify its driving style or provide soothing features such as music or lighting.

There are few papers that use ER in AVs for implemented applications, but several papers explored it without implementing an adaptive response. Several ER signal sensing methods have been explored for an AV application. Ma et al. used RGB cameras to observe facial expressions and head gestures inside a car [99], but their primary focus was to develop effective algorithms based on an existing dataset and did not consider the images in real driving scene. Furthermore,

[5]https://cordis.europa.eu/project/id/815003
[6]https://cordis.europa.eu/project/id/815001
[7]https://www.suaave.eu/

Table 2.3: Applying ER in existing and future HRI applications

	Application Areas	Methods	Emotion Model	Aim & Outcomes	Refs
EXISTING METHODS	Automated Vehicles	Facial expressions and head gestures	4 discrete emotions	Feelings of acceptance or satisfaction	[99]
	Automated Vehicles	Biophysiological signals	None	Feelings of acceptance or satisfaction	[96]
	Automated Vehicles	Speech and driving styles	None	Feelings of acceptance or satisfaction	[54]
	Social Robots (NAO)	Touch	8 discrete emotions	Robot design	[50]
	Social Robots (NAO)	Bodily movements	6 discrete emotions	Effective interaction	[51]
	Social Robots (NAO)	Facial expressions	7 discrete emotions	Better aid and support	[52, 177]
	Social Robots (NAO)	Multimodalities	Positive Negative and neutral	Older rehabilitation	[55]
	Social Robots (NAO)	Combination physiological signals	Positive, negative and neutral	Children impairment	[53]
	Social Robots (NAO)	Facial expressions physiological signals	2 dimensional emotions	Exercise monitor	[178]

	Possible Application Areas	Possible Methods	Potential Outcomes & Benefits
FUTURE METHODS	Remote Control (Hazardous & Inaccessible Environments)	Facial Expressions; Bodily Movements; Physiological Signals; Multimodal Data	Guarantee Security; Control Improve;
	Supervisory Control (Industrial Routine Tasks)	Facial Expressions; Bodily Movements; Physiological Signals; Multimodal Data	Human-centered; Balanced work; Social Acceptance; Trustworthy

biophysiological signals such as cardiac activity, electrodermal activity, skin temperature and respiration can be detected in cars [96], but these methods are influenced by the passengers' motions and need to be combined with other sensing methods to be effective. Speech can also be detected by microphones in cars, as can driving styles, grip strength, sitting postures, all of which can be used to facilitate ER [54]. Driver emotion is detected through a thermal camera [100], which is a non-intrusive car driver's emotion recognition, but the data was collected in a controlled environment. Overall, there are several potentially viable ER methods in AVs to acquire multimodal data from passengers, but many of these methods are at the preliminary stage and can suffer from interference from the passengers' motion artifacts and the in-car environments. As a result, these works are still at the speculative design stage and implementation and testing in-the-wild is lacking.

Emotion recognition for Human-robot social interaction

Several social robots are designed with social abilities to facilitate more natural interaction with humans, such as speech recognition and emotional intelligence [47]. Compared with the remote control, co-presence social robot interactions often aim to emulate human-like interactions [179]. For example, the robot may communicate with humans through visual, speech, and tactile communication. Based on this information, the robot can provide related and appropriate reactions or responses to assist or engage better with the user. This adaptive assistance approach which can be applied in sectors such as education, hotel service, therapeutic treatment, and entertainment [90–94]. Emotion intelligence (EI) is regarded as an essential ability in human interpersonal communications and even more critical than IQ [180], which is the ability to observe, interpret, generate, and express emotions. Endowing social robots with emotional intelligence will, thus, facilitate more natural and human-like communication. Existing ER work for HRI has primarily explored this context of this human-robot social interaction, which this section will now discuss.

NAO is a bipedal robot prominently used in this field. It can acquire touch signals, visual signals and audio signals to facilitate ER and EI. In [50], authors used human touch parameters, including intensity, duration, location and type of touch, on the NAO robot to recognize human emotions, highlighting the need for consideration of tactile sensor placement on the humanoid robots. Zhu et al. [181] developed a scheme for NAO robots to recognize emotions by analyzing human body visual information, which allowed the robot to perform natural interactive behaviors. Elfaramawy et al. [51] collected human full-body motion patterns from the depth camera attached to the head of NAO and used body skeletons to classify emotions, while Faria et al. [177] proposed a framework on NAO to recognise emotions using public dataset of facial expressions, although did not use emotional data produced in interactive social situations to train the model. In [52], authors integrated facial emotion recognition (FER) into NAO and the model was trained using the public FER dataset and tested using collected data using NAO. However,

the interaction was conducted in a controlled environment to ensure that the testing and training data were consistent. There are other challenges in facial expression recognition for HRI. Variations in lighting conditions, poses, and occlusions can affect recognition accuracy, as can variable distance between the human's face and the cameras attached to the robot, making the interaction limited to a certain range. In [44], they developed an algorithm that allows a robot to recognize human emotions throughout their non-stylized daily motions acquired by an attached Kinect sensor, to enable the robot to act in a more human-oriented way, although the datasets were not produced in real interaction scenarios. Existing work has explored using multimodal setups to capture user emotions during interactions with robots. In [55], authors augmented a socially assistive robot (SAR) using multiple signal input modalities, including facial expressions and speech, to recognize emotions of older adults during interaction. The experiments conducted on older adults living in care facility showed that empathetic SAR had positive effects on interaction and was more engaging and likable, however, this study included a small dataset and the dialogue was not open-ended. In addition to using sensors mounted on robots, sensors that detect the physiological signals can also be equipped for users to detect their emotions. In [53], they implemented ER in an assistive robot that supported children with hearing impairment, which could classify negative and positive emotions and showed that children had more positive emotions when interacting with the emotional robot. Participants wore a wristband to measure physiological signals, including skin conductivity and temperature, electrodermal activity, and BVP, and a video camera to collect facial expressions. In [178], authors implemented ER alongside a social robot for rehabilitation. The robot utilised used multiple signals, including facial expressions to detect user engagement, EEG signals to detect positive or negative emotions, and heart rate signals to monitor the physical activity during exercise and avoid excessive fatigue. The results showed that most participants has positive emotions during interacting.

2.2.6 Discussion on Future ER in HRI Applications

As robotics continues to evolve, gaining more capabilities and functionalities, their interaction with humans becomes more frequent and intimate. The future of Human-Robot Interaction (HRI) is unequivocally human-centred, which emphasizes the importance of considering human feelings, needs, behaviors and capabilities. The vision is for the robots to work alongside humans as partners, supporting human activities without taking over completely or making humans feel redundant. Furthermore, robots are likely to be perceived as more reliable and friendly, thereby enhancing people's willingness to incorporate them into their daily lives and workplaces.

To realize this vision, equipping robots with the capability to recognize and understand human emotions plays a critical role. Emotion recognition significantly enriches human-centred HRI in various ways, as previously discussed in Section 2.2.2, including allowing for more nuanced and context-aware communications, providing more personalized and satisfying interactions, increasing trust and acceptance and contributing to safer interactions in collaborative

environments. In the subsequent two subsections, we explore potential applications of emotion recognition (ER) in two underexplored areas of Human-Robot Interaction (HRI) including remote control and supervisory control of robots.

Future Trends of Emotion Recognition in Remote Robotic Control Scenarios

Human remote control of a robot involves a human operator controlling the robot's actions through a remote control interface. In this context, the operator sends commands to the robot through the interface, and the robot responds by performing the desired action [182], thus the remote robot can inherit human behaviors and is a physical avatar of the human. The human operator may be in a different physical location from the robot and may use various input devices, such as controllers (haptic device [183, 184]) and cameras (leap motion [185] and optitrack [186]) to send commands. Human remote control of a robot is essential in various applications, including space exploration [187], search and rescue operations [188], and military operations [189]. This type of control is often used when the robot is operating in a remote or hazardous environment, where direct human interaction is not possible or safe, to perform the mission-critical tasks [19, 182, 190]. The implementation of ER could further improve these systems, such as facilitating the prevention of danger or harm. For example, the emotions of a doctor may cause an imprecise operation while controlling a telerobot, causing severe harm to the patient. If the system can recognise problematic emotional states or behaviours and warn the operator or intervene, it could reduce adverse impacts on the remote environment. ER could also be used to recognise emotive features in controlled-robot movements, which could be dynamically normalised in real time to help the operator maintain smooth and safe control of using shared control [191]. No ER work in remote control exists, despite these potential upsides.

Future Trends of Emotion Recognition for Human Supervisory Control of Robots

The fourth category of HRI application is the human supervisory control of a robot to conduct industrial tasks, such as pick-and-place, assembling and delivery. During this process, humans and robots need to cooperate to complete the tasks in the shared workspace. Human workers monitor and direct the autonomous or semi-autonomous actions of one or more robots, rather than taking direct control. The human operator has a higher-level understanding of the robot's task and provides high-level commands and goals to the robot. The robots adapt to workers' behaviors and changing circumstances to be in conjunction with workers, which are regarded as collaborative robots or *cobots* [179]. The fifth Industrial Revolution (Industry 5.0) aims to be human-centered, and the collaboration between human and cobots is one of the keywords in Industry 5.0 (proposed by the European Commission) [192]. Compared with the remote control of robots, a human supervisor is not directly involved in robot interaction, instead they monitor the robot and thus their emotions do not directly influence the robot's operation. However, their emotions can still reflect their interest and satisfaction with the work, which could be used

to configure a balanced work proportion between humans and robots, or identify persistent or concerning issues with current operational procedures.

During cooperation, emotion is an essential but often overlooked factor. In [193], authors introduced the challenges of cobots, including social acceptance, security, and cognition. One aspect of social acceptance is the human response to robots' characteristics or actions, which ER can be used to analyze. Furthermore, safety and security issues can be addressed by checking the nervous, anxious, and annoyed emotions to avoid a negative impact on the operation. Lu et al. [49] indicated that there are five levels of industrial human needs. The first level is safety: robots must predict and adapt workers' behaviors. The second level is health: robots should anticipate workers' physical fatigue and mental health to minimize the impact on the workers' health. The third level is belonging: robots should respond to the workers' social attributes and to human emotional needs, building trustworthy human-robot relationships and enabling them to cooperate with common goals via mutual communication and understanding. The fourth level is self-actualization: in which humans achieve self-fulfillment as the system offers personalized co-learning, enabling bi-directional learning between human and machine agents. Based on these, in Industry 5.0, human emotion recognition is a crucial component. For example, emotions are used to evaluate human physical fatigue, allowing the robot to make relative judgements, such as taking on more work. Emotions can also be regarded as mental health cues or can be used to evaluate human satisfaction with robots, facilitating present or future improvement of the user's experience with the robot. In addition, understanding human workers' emotions allows industrial robots to exhibit emotional intelligence and adapt to their coworkers' emotional needs, taking this field to new empathetic heights. We now discuss how existing methods of ER could be applied to this area in future work and the existing challenges.

ER Challenges in HRI Scenarios

Choosing how to conduct ER in these scenarios could be a challenge. Physiological signals can reveal a human's instinctive affective states [194], but there are some limitations when using one of these modalities recognize emotions in remote control applications. Some peripheral physiological data need humans to stay stable, such as ECG, EEG [60] and HRV [195], while others, such as SKC and GSR, can suffer from latency issues [60]. In remote control, the human operator moves while controlling, so extracting implicit physiological signals also provide a challenge as motions can cause artifacts in the electrical signals [159,196]. Alternatively, FER can be used, but most existing methods have used public data and were not implemented in real interaction scenarios, where some problems could occur, such as uncontrolled environments, obstacles, lightning and the user's moving face. Cameras can also be used to observe bodily expressions, but they face similar problems to FER caused by the camera-captured data. Methods using a single part of the body, such as hands, arms and heads, need further exploration, as operators do not move their whole body during control in many use cases. Other options, such as voice,

text, and handwriting signals, do not necessarily appear in remote-control scenarios. Based on these, the combination of different signals may be a solution to improve the performance in remote control scenarios. Apart from building multimodal systems, emotion recognition methods primarily focus on algorithmic research and experimental environments, so enhancing the generalization of algorithms is essential. For example, using data augmentation to increase the diversity of the data and using regularization to reduce overfitting. Additionally, applying motion artifacts removal methods [197] in physiological signals could allow them to be used in HRI scenarios. Furthermore, improved human detection methods of visual information could promote facial and bodily signals applied to classify emotions in real HRI scenarios.

2.2.7 Conclusions

Emotion recognition is an irreplaceable ability in interpersonal communication. We can predict other people's motivations by understanding emotions. Up to now, robots mainly perform industrial tasks without much consideration for the emotions of the humans they share interaction or space with. However, the final aim of robots is to serve human beings as effectively as possible. Based on this, it is increasingly important to endow robots with emotional intelligence. In this work, we comprehensively review the existing literature on ER in HRI, existing HRI application types and state-of-art ER work in HRI. The most effective ER methods for social robots are facial, voice and bodily expressions, which are interpersonal information. For AVs, it is more effective to use multimodal physiological data, as well as driving telemetry. In remote control scenarios, it is more effective to use human operation data, such as arm movements. In supervisory control of robot applications, facial and physiological data are promising, although further exploration is needed. This work provides an overview of the whole field of ER and its intersection with HRI, as well as key takeaways for which ER sensing methods may be appropriate for different applications and which current and future applications are promising. We also give our view on the most pressing under-explored areas of ER in HRI, including remote control and supervisory control. We further provide practical ER methods for these two HRI applications to encourage future research. In conclusion, we aim to promote a future-facing approach to promote the next generation of emotionally intelligent robotics.

2.3 Literature Review Conclusions

Teleoperated robotic system is vulnerable to cyber and physical attacks. Encryption can protect the cyber attack during communication, but it adds latency and decrease the control efficiency of the teleoperated robot system. Robot misbehavior detection method can protect the robot device, but it is applied in the autonomous robot system and can not protect the spoofing. User authentication method can protect the access point of the system, but it can not guarantee the security of the signals being tampered during long-distance transmission.

Emotion factors can also influence the robotic system's security. The existing emotion signals are physiological signals, facial signals, semantic signals and body signals. We studied the emotion recognition method in four crucial application areas of HRI [19] including human supervisory control of robots, human remote control of robot, automated vehicles and human–robot social interaction with robot. We demonstrated the overall state-of-the-art techniques in most HRI related fields, including eye tracking and the use of new AI techniques in the applications of ER. Additionally, we for the first time discussed the potential challenges and opportunities of emotion recognition in the area of HRI in the future.

Chapter 3

User Classification from Motion-Controlled Robotic Behaviors

3.1 Introduction

Common methods of robotic security include abnormal signal detection, password-based login and user classification using behavioral biometrics. Password-based login allows users to enter pre-set passwords to the master controller platform. Behavioral biometrics identify focuses on the unique patterns in human behaviors that are unique and difficult to replicate or forge, which could provide a continuous user classification. However, these two methods have a key limitation: they can not guarantee the data does not tamper when transmitting from user master controllers to the remote slave robot, which causes the remote robot misbehavior that may damage the local environments and loss human operators' trust. Based on this, using the robotic beahviors to do the user classification can guarantee the data is not tampered with until the robot is executed, which was discussed in this Chapter.

The first experiment built a motion-controlled Franka robotic arm. The robot followed human operators' trajectories in real-time. We verified that the robotic arm can inherit human biometric identities through the trajectories and user classification accuracy reached 95%. This experiment provided a foundation for the next three experiments, which continuously investigated the user classification using robotic behaviors. Experiment 2 implemented user classification on motion-controlled social robot trajectory. Although the identity information in the behaviors of one arm was slightly lost, the identity information from two hands of the NAO robot increased and the user classification result of two arms' end-effectory trajectories reached to comparable results with one Franka arm. Besides, users can be classified through social tasks, which provides the foundation of the real application. Experiment 3 investigated three controlling parameters' influences on robotic behaviors. When we adjusted these parameters to decrease the system performance, the user classification was decreased. We found that lower system performance leads to a lower user classification performance. However, allowing remote

areas to know the user's identity leaks user privacy to the remote areas. In experiment 4, we implemented Reinforcement Learning (RL) on the captured operators' trajectory data to keep the operators' privacy.

3.2 Background

3.2.1 Robotics Control

Robotic Kinematics

A robotic manipulator consists of a series of rigid links connected in a serial configuration, where each link is joined to the next by a joint [198]. The common joint types are revolute, prismatic, and continuous joints. Besides, a robot's degree of freedom (DOF) is the total number of freedoms minus the number of constraints. Motors are employed to control the movement of these joints by transforming electrical energy into mechanical motion. The movement of the joint drives the movement of the links, eventually driving the movement of the effector at the end of the robotic manipulator, known as the robotic end-effector. The end-effector is the component designed to interact with the environment, such as grasping objects, walking, shaking heads or performing other tasks. In robot kinematics, forward kinematics is to apply the robot's kinematic equations to determine the end-effector's position based on given joint parameter values [199]. On the other hand, inverse kinematics focuses on determining the values of the robot's joints based on the position and orientation of the end-effector.

Robot Control Methods

The robotic controller receives the sensor signals and then controls the actuators to achieve the target behaviors of the robot. Robotic control mainly has four categorizations including motion control, force control, hybrid motion-force control, and impedance control. Motion control with velocity inputs is a common method, known as velocity control, to allow robots to achieve the required velocity. The proportional integral derivative (PID) control algorithm that is stable and robust control of systems is widely used to achieve velocity control. The PID controller includes three components proportional where output is proportional to the error, integral where output is adjusted in regard to the sum of errors, and derivative where output is adjusted by the change rate of the error. They are combined together to maintain the desired output, shown as 3.1. In 3.1, *e(t)* is the error at time t and *u(t)* is the control output at time t. The terms K_p, K_i and K_d correspond to the proportional, integral, and derivative gains respectively.

$$u(t) = k_p e(t) + k_i \int_0^t e(t)dt + k_d \frac{\mathrm{d}e(t)}{\mathrm{d}t} \tag{3.1}$$

3.2.2 AI for Robtics

Artificial Intelligence (AI) has profoundly revolutionized numerous fields, such as healthcare, financial services, manufacturing, education and robotics. The integration of AI into robotics has resulted in enhanced capabilities in robots, enabling them to perform complex industrial tasks, adapt to their environments, and work alongside humans. With the robotic endowing more and more capabilities, they have more work in different scenarios. They have more opportunities to interact with people and work more frequently and closely with human. There are several commonly used AI techniques applied in Human-Robot Interaction (HRI), including Convolutional Neural Networks (CNNs), Dynamic Time Warping (DTW), K-Nearest Neighbors (KNN) and Support Vector Machine (SVM). CNNs have been widely applied in various HRI scenarios. For instance, in face-to-face interactions between a social robot and a human, a robot equipped with a CNN can interpret the information conveyed by the human through semantic analysis. DTW is used for the analysis of human movement time-series data to classify emotions. SVM is applied for the analysis of ECG data to detect human heart health, mental health and emotion.

CNN

Classification tasks are fundamental challenges that neural networks are designed to address. The structure of neural networks are simulations and simplifications inspired by biological neurons found in human and animals. Multi-layer Perceptron (MLP), regarded as the conventional fully connected (FC) networks, is the fundamental form of neural network and contains three layers including input layers, hidden layers and output layers, shown as 3.1. The forward propagation of data between neurons is by using values weights and bias. For example, the input *X1* forward propagates to *H1* neuron by calculating weight (*W1*) multiply input (*X1*) plus bias (*b1*) seen in 3.2. After calculating and adding together all the input to this neuron, the value in this neuron passes through an activation function f such as ReLU, Sigmoid and Tanh to get the output with the non-linear feature. After forward propagation and getting a prediction value, the network's prediction value is compared to the actual target value to calculate the cost function, or loss function, such as square error and cross entropy. Finally, the optimization function is used to minimize this error by adjusting the weights and biases. The network undergoes multiple iterations of forward and backpropagation on the training data, adjusting its weights and biases to minimize error. After sufficient training, the network's predictions should match the actual values, indicating that it has learned the underlying patterns in the data.

$$H1 = W1 * X1 + b1. \tag{3.2}$$

However, when the input data is large, FC networks produce abundant parameters. Compared with conventional FC networks, CNN utilize shared weights and local connections to effectively leverage the two-dimensional structure of input data, such as images [200]. Based

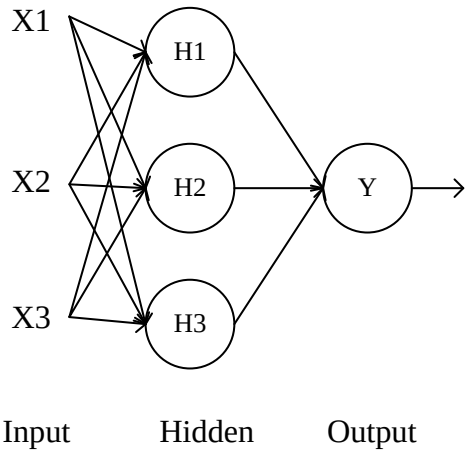


Figure 3.1: Multi-layer Perceptron (MLP) Structure

on this, this operation not only simplifies the training process but also accelerates the network's performance, by employing a significantly reduced number of parameters [200]. CNNs are particularly powerful for tasks related to image recognition, classification, and analysis. CNNs generally contain 5 types of network structure including an input layer, convolutional layer, activation layer, pooling layer, fully connected layer and output layer. In the input layer, the raw data is pre-processed by using methods, such as normalization, regulazation, PCA and whitening. Then, in the convolutional layer which is the core block of CNNs structure, features are extracted by the kernels, or filters, multiplying with the input matrix. Suppose we have one $N \times N$ matrix as input defined as I and one $m \times m$ matrix as kernel defined as K. After the convolutional layer, the convolutional layer output will be the size of $(N-m+1) \times (N-m+1)$. The convolutional calculating equation is defined as 3.3, where I(i, j) is the value of I at position (i,j) and K(m, n) is the weight value of the kernel at the position (m, n).

$$(I * K)(i, j) = \sum_{n} \sum_{m} I(i+m, j+n) \cdot K(m, n) \tag{3.3}$$

The activation layer is applied to introduce non-linear properties to the system. The pooling layer, also called the down-sampling or subsampling layer, is often used after the convolutional layer and is used to reduce the feature dimension of the convolutional layer output, which can effectively reduce network parameters and prevent overfitting. Common pooling methods include general pooling, overlapping pooling and spatial pyramid pooling. The fully connected Layer is responsible for summarizing the features extracted from the convolutional and pooling layers and mapping the multi-dimensional feature input into a two-dimensional feature output. The architecture of one CNN is shown as 3.2. CNN has applied many fields, such as image processing [201], natural language processing [202] and speech processing [203].

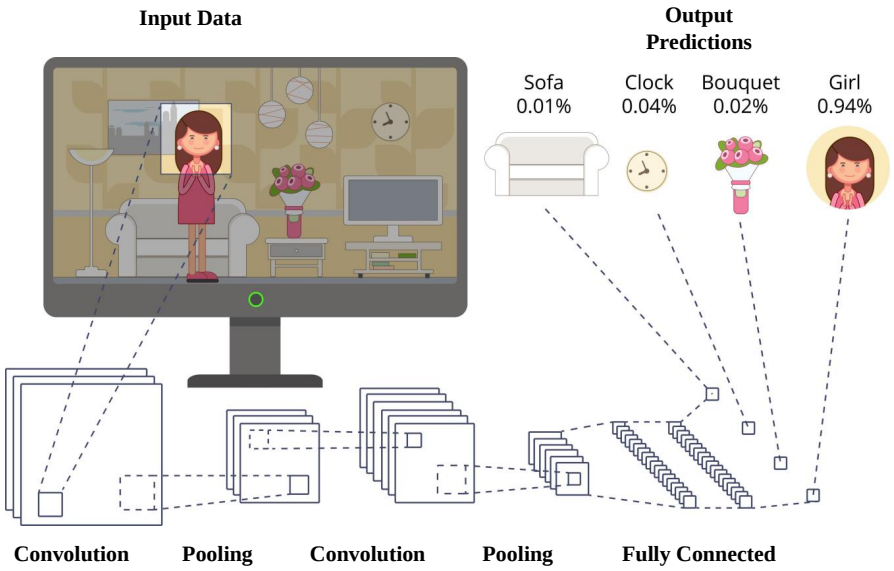


Figure 3.2: CNN Structure

Then, the training set of input x and output y are fed into the network model. After, training the model and learning the parameters. However, the model may overfit the training data, so in order to limit the overfitting, validation data is used for evaluating the model and adjusting the hyperparameters, such as learning rate, epoch number, batch size and dropout. The common methods of validation include hold-out validation and cross validation. Hold-out validation provides a static partition, splitting the dataset into training, testing and validation datasets randomly. However, when the dataset is small, a random partition makes different results for different test sets. To deal with this, cross-validation is implemented, which repeatedly partitions the dataset into train and test sets. The common methods are K-fold and Leave-one-out. K-fold cross-validation is to split data into K groups one of which is used for training and the rest is used for testing. Leave-one-out cross-validation (LOOCV) is when K equals to one. In LOOCV, a single data sample is selected as the test set, while the remaining samples are used to train the model each time.

Finally, the performance metrics are used on the results to evaluate the model. For the classification model, the common metrics are confusion matrix, accuracy, precision, recall and F1 Score. The confusion matrix is composed of the true positive (TP) which is the number of positive samples that are classified correctly, false positive (FP) which is the number of negative samples that are classified incorrectly, true negative (TN) which is the number of negative samples that are classified correctly, and false negative (FN) which is the number of positive samples that are classified incorrectly. Classification accuracy is calculated by dividing the number of correct predictions by the total number of predictions and then multiplying the result by 100 to

express it as a percentage.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{3.4}$$

Precision is calculated as the ratio of true positive samples to the total number of predicted positive samples.

$$Precision = \frac{TP}{TP+FP} \tag{3.5}$$

Recall is fundamentally the proportion of true positive sample relative to the total number of actual positive samples in the ground truth data.

$$Recall = \frac{TP}{TP+FN} \tag{3.6}$$

The F1-score metric uses a combination of precision and recall. In fact, the F1 score is the harmonic mean of the two. The formula of the two essentially is:

$$F_1 = 2 \times \frac{precision \times recall}{precision + recall} \tag{3.7}$$

DTW

It is changeable to compare the similarities of different temporal sequences, given the fact that in practical scenarios two different time series exhibit approximate similarities along the time axis, but their exact correspondences remain unclear. For instance, when two individuals utter the same word, variations in their voice timbre and frequency mean that correspondence of the two voice data between specific moments may not align while the pronunciations sound similar. Thus, the DTW algorithm stands out and shows high performance in understanding the alignment and similarity between time-sequenced data. DTW has the capability to stretch and compress segments of the data to find an optimal match between two time series sequences.

DTW is essentially a dynamic programming algorithm. For example, we have two time series sequences. One is denoted as $f_q = f_q(1), f_q(2), f_q(3), ...f_q(i)..., f_q(N)$, where N is the time length. The other one is denoted as $f_w = f_w(1), f_w(2), f_w(3), ...f_w(j)..., f_w(M)$, where M is the time length. When N and M are equal. We can directly calculate the distance of them using Euclidean distance. Otherwise, we need to find the shortest distance between them by building a $N \times M$ matrix. The The distance between point $f_q(i)$ and $f_w(j)$ is calculated as $d(f_q(i), f_w(j)) = (f_q(i) - f_w(j))^2$. The DTW algorithm seeks to find the shortest path from the origin to $(f_q(n), f_w(m))$. We define the path as 3.8 and DTW is calculated as 3.9.

$$W = w_1, w_2, w_3, ..., w_k \quad max(m,n) \leq K < m+n-1 \tag{3.8}$$

$$DTW(f_q.f_w) = min\left\{\sqrt{\sum_{k}^{k=1} w_k/k}\right\} \tag{3.9}$$

KNN

The KNN algorithm is a widely utilized machine learning technique for tackling classification tasks. Its fundamental principle involves determining the similarity between training data and test data based on distance metrics. The KNN classifier performance relies on a distance metric. The higher the ability of that metric to find the similarity between data, the higher performance the classifier will have. The most commonly used distance measure is Euclidean distance 3.2.2. In practice, KNN identifies the k closest training examples to a given test point and then classifies this point based on the majority class among these neighbors.

$$d(x,y) = \sqrt{\sum_{i=1}^{m}(x_i - y_i)^2} \tag{3.10}$$

3.3 Experimental 1: User Classification of Motion-Controlled Franka Robotic Arm

The first experiment was to classify different human operators through the motion-controlled robotic arm's behaviors. It was driven by three primary objectives. We built a motion-controlled robot platform utilizing a Franka robotic arm. Two gestural motions with significant real-world applicability were designed, reflecting natural methods of human communication and control in HRI contexts. The introduced tasks were the "in-air" task and the "line-tracing" task, each rooted in practical HRI scenarios. The in-air task finds its application in contexts such as remote robot education [204] and telehealth rehabilitation on robotic platforms [205], while the line-tracing task is crucial in mission-critical environments like remote robotic dentistry [206] and robot-assisted spine surgery [207]. Subsequently, we applied two algorithms to distinguish among 10 different users across these tasks, achieving a classification accuracy of 95.0%.

Therefore, the first aim was to validate that the motion-controlled robotic arm could inherit human behaviors and that users could be uniquely identified through robotic data. Secondly, to demonstrate the application of this methodology to tasks with genuine real-world relevance, laying the groundwork for its method in remote control robot scenarios. To provide a preliminary testing ground for task design, experimental setup, data collection methods, and algorithm implementation, all of which are poised for future research endeavors.

3.3.1 System Overview

The aim of this system is to verify that the motion-controlled robotic arm can inherit human operators' behaviors and that three different implemented algorithms can identify users through robotic joint data. The system overview is illustrated in Figure 3.3, which mainly includes two parts. One is the user end where users' motions are captured and control commands are generated. The other is the robotic arm end for executing control signals and generating motions where the learning-based user authentication happens. The green box shows the human operator. An operator wore hand gloves and attached three markers to control the robot to perform motion tasks in the air. Then the cameras captured the trajectory of the markers. Then, the acquired position and orientation of human hand trajectory are converted to 7 joint angles of the robot with inverse kinematics. Next, we provided velocity control commands to robot joints to move the end-effector to the desired position and orientation. Meanwhile, operators monitored the movements of the robotic avatar and modifies their own actions to facilitate interactive control through hand-eye coordination. Two user classification algorithms were deployed on the robotic arm's joint to classify different users.

3.3.2 Motion-Controlled Franka Robotic Arm Platform

We built a motion-controlled Franka Emika Panda Robot [208] arm platform, where a human interactively controled a robot arm, as shown in Figure 3.4. Franka is a state-of-the-art robotic arm, which has seven joints (7 DoF) and provides high-precision performance. The motion capture system is OptiTrack [209], which is designed to deliver sub-millimeter accuracy in tracking motion and provides real-time feedback. These six cameras were arranged in a circle to capture the operator's hand motion trajectory via a glove attached with four markers. There were two personal computers (PCs).

The first PC calculated the position and orientation of human hand trajectory in a Cartesian coordinate system and sampled the data at 120 Hz. Then, these data were sent to a second PC using a local network. The second PC provided velocity control using PID path-planning 3.2.1, in which joint angle values were then converted into robot adaption commands. Finally, the robotic end-effector moved to the desired position and orientation. When the robot received commands and performed motions, users observed the robot's behaviors and adjusted their own behaviors to interact with the robot forming a control loop.

3.3.3 User Classification Algorithm Design

In this section, we introduced the method of robotic arm end-effector reconstruction and gave descriptions of the methods for user identification.

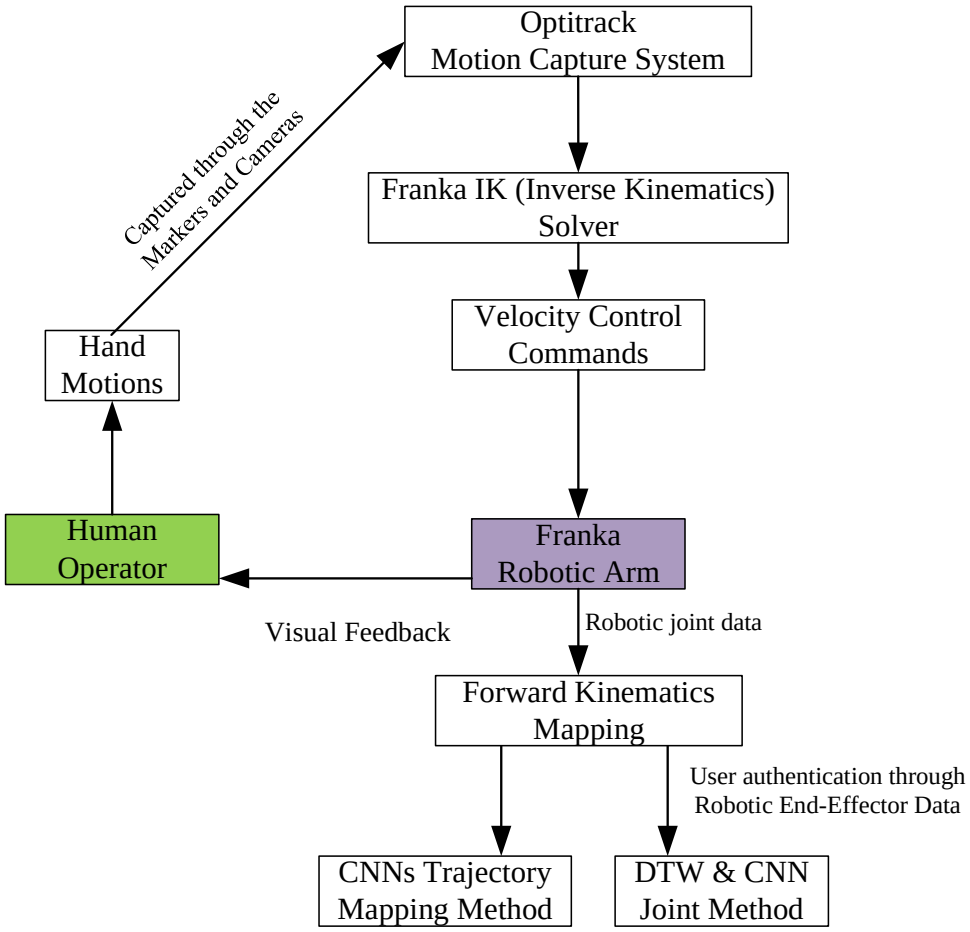


Figure 3.3: The overview of user identification through motion-controlled Franka robotic arm system.

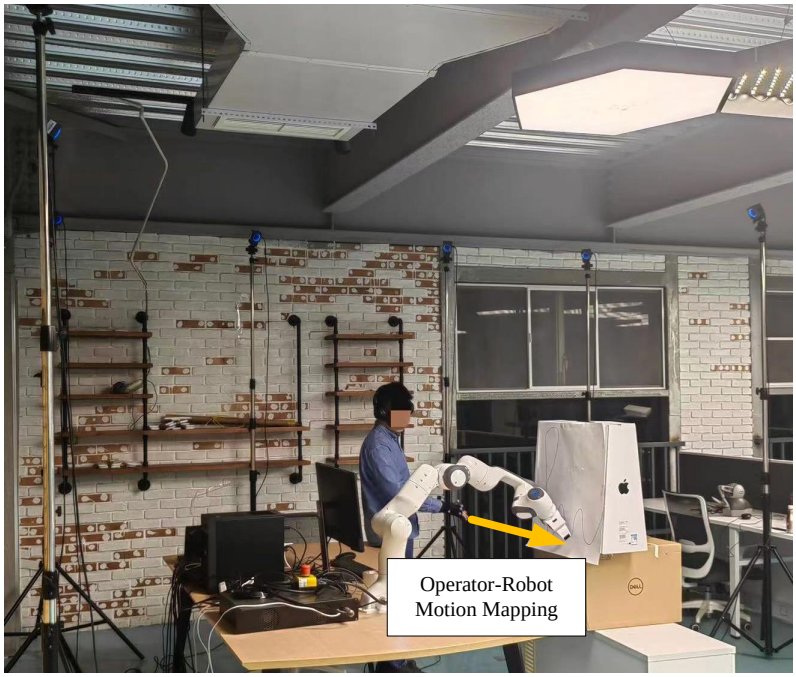


Figure 3.4: The platform of user identification from motion-controlled Franka robotic arm.

Franka Robotic Arm Forward Kinematics

Forward kinematics is a fundamental concept in robotics that involves determining the position and orientation of the robot's end-effector given the values of its joint variables. Denavit-Hartenberg method is the most common method for robot forward kinematics [198], allowing engineers and roboticists to predict the final location of the end-effector based on the parameters of each joint. There are four parameters of the Denavit-Hartenberg method including a_i, d_i, α_i and θ_i, which are the name of link twist, link offset, link length and joint angle, respectively. There is the coordinate frame for each link aligned by DH parameters. ${}^{i-1}T_i$ is the transformation matrix which describes the position and orientation of frame i in frame i-1.

$$ {}^{i-1}T_i = \begin{bmatrix} \cos\theta_i & -\sin\theta_i & 0 & a_{i-1} \\ \cos a_{i-1}\sin\theta_i & \cos a_{i-1}\cos\theta_i & -\sin a_{i-1} & -\sin a_{i-1} d_i \\ \sin a_{i-1}\sin\theta_i & \cos\theta_i \sin a_{i-1} & \cos a_{i-1} & \cos a_{i-1} d_i \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{3.11} $$

The kinematic chains are assembled by links and joints to produce a desired motion. 3.12 calculates a kinematic chain of 7 DOFs manipulator by multiplying all the transformation matrix starting from ${}^0T_1$ to ${}^6T_7$. ${}^0T_7$ describes the position and orientation of the end-effector frame in the base frame.

$$ {}^0T_7 = {}^0T_1 * {}^1T_2 * {}^2T_3 * {}^3T_4 * {}^4T_5 * {}^5T_6 * {}^6T_7 \tag{3.12} $$

The transformation matrix is composed by a rotation matrix and a translation matrix shown as 3.13. $\hat{R}$ defines the rotation matrix and the $[d_x d_y d_z]$ defines the translation.

$$T = \begin{bmatrix} & \widehat{R} & & \begin{bmatrix} d_x \\ d_y \\ d_z \end{bmatrix} \\ \begin{bmatrix} 0 & 0 & 0 \end{bmatrix} & & & 1 \end{bmatrix} \tag{3.13}$$

Based on this, we calculated the 3D coordinate position of the end-effector by multiplying ${}^0T_7$ and [0 0 0 1]. After this, we reconstructed the robotic end-effector data by using the 7 joint values.

Data Segmentation and Normalization

After determining the values for the end-effector, we segmented the data into distinct instances based on timestamps to identify the start and end points of each instance. In our study, an instance referred to a single task trajectory for a specific operation, or we considered it a single instance each time a user instructed the robot to perform a particular task trajectory, thus an instance was a 3D trajectory. Subsequently, we normalized each instance as outlined in 3.14, scaling them within the range [0, 1] to make each instance comparable.

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}} \tag{3.14}$$

CNN Algorithm Design

The end-effector data is 3D trajectory data, as shown in Figure 3.7, which contains both time and spatial information. We utilized two steps to encode the spatiotemporal information of 3D images into 2D images. Wang et al. [210] proposed an approach for encoding the spatiotemporal information contained in 3D skeleton sequences into multiple 2D images. Based on this, we first converted the 3D image into three 2D grayscale images, during which the 3D end-effector trajectory is projected to three viewings including the front plane, top plane and side plane of the 3D image, and the spatial information of the end-effector trajectory was preserved. Secondly, we encoded the time information into the 2D image and the temporal information is shown by the color gradient approach, using the color brightness change to present how fast the end-effector is moving. An end-effector trajectory-based representation of robot motion is shown in Figure 3.5. The three greyscale images are regarded as three channels of RGB images and then fed into a CNN-based model for user classification. The CNN contains three convolutional layers and is followed by a fully connected layer and a dropout layer to avoid overfitting.

(a) Front. (b) Side. (c) Top.

Figure 3.5: The three-view of end-effector trajectory.

KNN+DTW

Feature Extraction:

We first extracted 11 features related to human behavioural biometrics from robotic end-effector trajectory. These kinematic features are commonly used biometric features extracted from human motion for user identification analysis. These biometric features are used on different human behaviors including gait [211], hand movement [138], gestures [212] and mouse dynamics [213]. In specific, 3D position over time provides a detailed understanding of movements. Velocity, which refers to the speed and direction of movement, can help differentiate between similar actions performed at different speeds. Acceleration which is the speed of a movement that changes over time is another biometric feature. Different individuals may accelerate or decelerate their movements in unique ways, even if the overall pattern looks similar.

These 11 features are time series and donated as vectors $p(t), \dot{p(t)}, ||p(t+1)-p(t)||, \ddot{p(t)}$, $||\ddot{p(t)}||$, where the t means the time sampling index, t=1,...,N. We illustrate the mathematical formulas of biometric features analyzed in our study and how they are constructed below.

Feature Construction:

- *Vector of x-axis of trajectory:* $\mathbf{X}$(t)
- *Vector of y-axis of trajectory:* $\mathbf{Y}$(t)
- *Vector of z-axis of trajectory:* $\mathbf{Z}$(t)
- *Vector of x-axis of velocity:* $\dot{\mathbf{X}(t)}$
- *Vector of y-axis of velocity:* $\dot{\mathbf{Y}(t)}$
- *Vector of z-axis of velocity:* $\dot{\mathbf{Z}(t)}$
- *Vector of x-axis of acceleration:* $\ddot{\mathbf{X}(t)}$
- *Vector of y-axis of acceleration:* $\ddot{\mathbf{Y}(t)}$
- *Vector of z-axis of acceleration:* $\ddot{\mathbf{Z}(t)}$
- *Vector of Euclidean norm of position frames difference:* ||$\mathbf{P}$(t+1)-$\mathbf{P}$(t)||

- *Vector of Euclidean norm of acceleration difference:* $\|\mathbf{P}(\ddot{t+1}) - \mathbf{P}(\ddot{t})\|$

As the feature vectors shown above, 3D trajectory is constructed by vector of **P**(t)=[**X**(t), **Y**(t), **Z**(t)]. 3D velocity is constructed by vector of $\mathbf{P}(\dot{t})$=[$\mathbf{X}(\dot{t})$, $\mathbf{Y}(\dot{t})$, $\mathbf{Z}(\dot{t})$]. 3D acceleration is constructed by a vector of $\mathbf{P}(\ddot{t})$=[$\mathbf{X}(\ddot{t})$, $\mathbf{Y}(\ddot{t})$, $\mathbf{Z}(\ddot{t})$]. The position difference between time frames is to calculate the straight-line distance between two points in 3D space using Euclidean distance. The acceleration magnitude is to calculate the Euclidean distance of three-axis values of acceleration. Finally, we construct all these 11 vectors into a matrix as one instance, *I(t)* = $[p(t), p(\dot{t}), ||p(t+1) - p(t)||, p(\ddot{t}), ||p(\ddot{t})||]$.

Feature Normalization:

After extracting these 11 features, we applied feature scaling, also called feature normalization, to standardize features. This technique involves rescaling each single feature to the range of 0 to 1. This adjustment is particularly important for machine learning algorithms whose features vary significantly in scale due to the diverse properties they represent. Thus, standardizing makes each feature contribute equally to the classification analysis.

KNN+DTW algorithm:

KNN, as detailed in Section 3.2.2, identifies similarities between data points by computing their distance. However, for time series data, DTW provides a more efficient way to calculate distances than the traditional Euclidean distance. As discussed in Section 3.2.2, DTW is particularly suitable for analyzing temporal signals where the sequences are not aligned. Thus, the combination of KNN and DTW allows KNN to more accurately classify time series data, as demonstrated in the work by Marcacini et al. [214]. We used the DTW to find the similarities between testing and training data. In specific, we split 60% of the dataset into training sets, while the remaining 40% was the testing sets. For each test instance, we compared the distance between it and the training instance and found the nearest 5 training instances as its predicted class. Figure 3.6 illustrates the details of the KNN+DTW algorithm, where the green box is the testing instance, while the green and blue circles are training instances with different labelled user classes.

3.3.4 Experiments

In this section, we described the experiments of user classification from the motion-controlled Franka robotic arm platform, including experiment setup, task design and data collection.

We recruited ten different volunteers to participate in this experiment. Half of them are females and remain are males, all of whom are students from college and the ages range from 19 to 22 (mean=20.2, σ =1.2). All these ten volunteers signed the consent forms and were given 5 minutes to be familiar with the robotic controlling before data collection. Then, they were given 5 minutes to control the robot, by performing some random tasks in the air while observing the robotic behaviours and adjusting their controlling behaviors based on the robotic feedback.

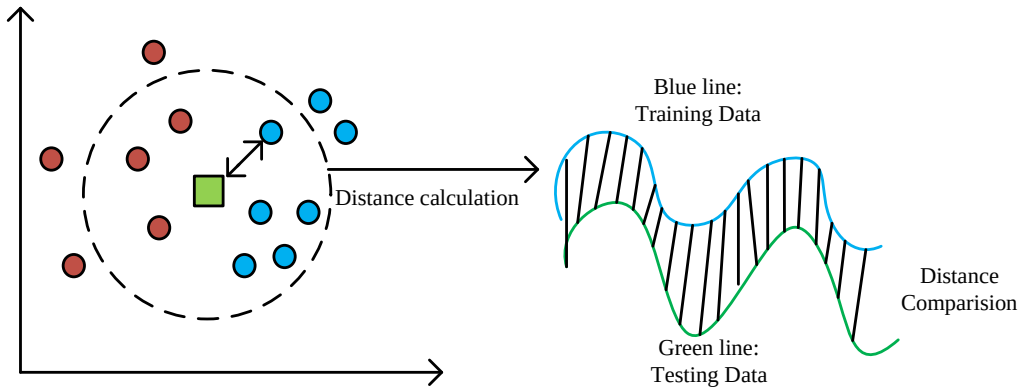


Figure 3.6: The DTW and KNN combination Algorithm

We developed two types of tasks for this study: in-air and line-tracing tasks. The in-air task involved participants writing the initials "LW" [215] in the air, employing their writing style. For the line-tracing task, participants were instructed to trace a predefined version of "LW," as illustrated in Figure 3.4. The initials "LW" stand for "Lucky Win." Each participant executed the robotic arm to perform 30 times for each task. As we mentioned in Section 3.3.3, each time repetition represents one instance. Consequently, the total number of instances generated was 6000 (10users$\times$ 30 samples$\times$ 2 tasks).

3.3.5 Results Analysis

We verified that the robotic can inherit human behaviors and evaluated our user classification method based on the motion-controlled Franka Emika Panda Robot platform. Besides, we compared the results on robot and human data respectively.

The Similarity between Optitrack Captured Human Arm and Franka Robotic Arm End-Effector Trajectory

We plotted the 3D motion of the human arm and robotic arm in Figure 3.7. The read trajectory is the human hand trajectory captured by the Optitrack, while the blue one is the robotic arm's trajectory motion controlled by a human in real time. This figure shows that the robot can show a similar trajectory as the human. The difference between them is caused by the precision difference between the robot and the motion capture system. The Optitrack, as a high-precision motion camera, is able to demonstrate a more accurate human movement, while there could loss of some information on the robot.

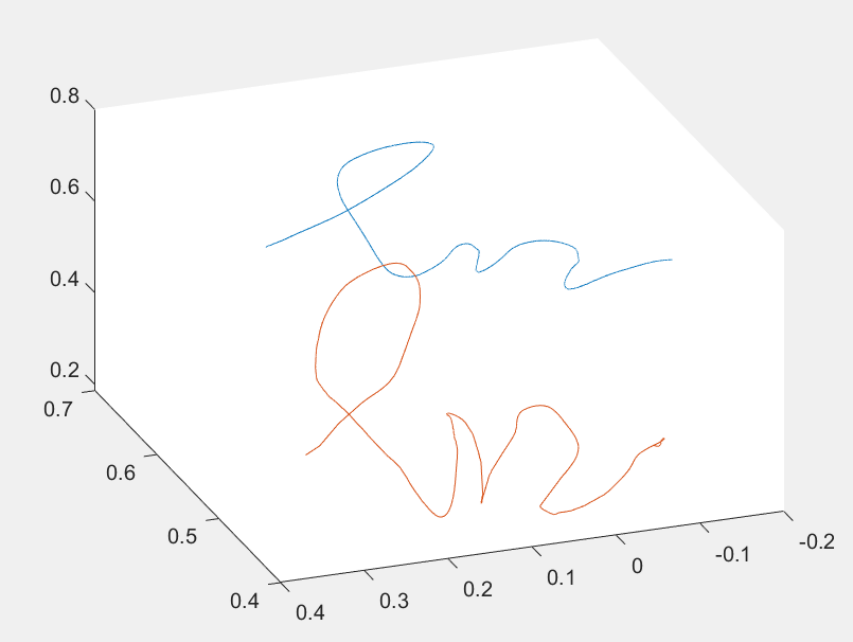


Figure 3.7: Human arm trajectory (red) VS robotic arm trajectory (blue).

Table 3.1: The user classification accuracy under the combination of task types and data types.

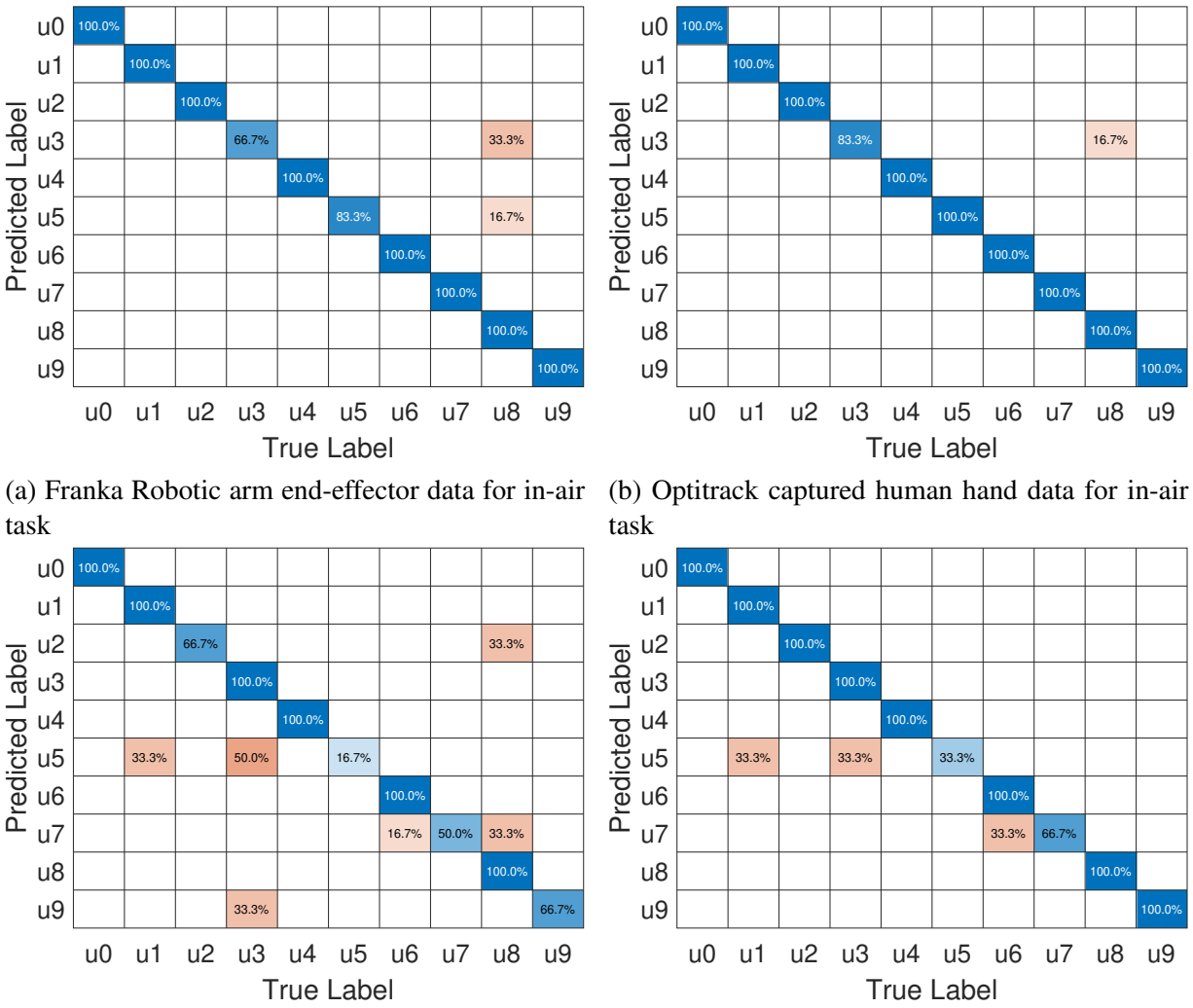
	Franka robotic arm end-effector data	Optitrack captured human hand data
Line-tracing task	0.8	0.9
In-air task	0.95	0.98

KNN+DTW User Classification Results

Table 3.1 shows that the highest user classification accuracy of using robotic arm data reaches 95.0%. The accuracy of using robotic line-tracing task data is lower, that is 80% compared with that of using in-air data. This is reasonable because the predefined reference trajectory reduces the 3D position variance among users. Consequently, this underscores the efficacy of higher-level features, such as velocity and acceleration, in differentiating between users. In addition, the human data shows a similar conclusion on the in-air and line-tracing tasks.

Besides, we compared our user classification results on the robotic data and human data, respectively. In Table 3.1, we can observe that for both tasks, the classification accuracy is higher with human trajectory data than with data from the robotic arm. This supports the insights drawn in Section 3.3.5, indicating that while the motion-controlled robotic arm can capture much of the human behavioral information, there is a loss of certain features that leads to lower classification performance for the robotic arm trajectories.

Furthermore, we presented each type combination's confusion matrix in Figure 3.8. For the in-air task of both the robot and line-tracing task, users 0, 1, 4, 7, 8, and 9 are classified with 100% accuracy, while there is a confusion between user 3 and 8, shown as Figure 3.8a and Figure 3.8b. There is a lower performance of the line-tracing task for 10 users on both robot



(a) Franka Robotic arm end-effector data for in-air task

(b) Optitrack captured human hand data for in-air task

(c) Franka Robotic arm end-effector data for line-tracing task

(d) Optitrack captured human hand data for line-tracing task

Figure 3.8: The confusion matrix of user classification for in-air and line-tracing tasks using Optitack captured data and robotic end-effector data, respectively.

and human data, shown in Figure 3.8d and Figure 3.8c. We can see that for each user there is a performance decrease on the line-tracing task.

CNN User Classification Results

The performance metrics of the CNN model for user classification based on robotic arm end-effector data are presented in Tables 3.2 and 3.3. Specifically, the accuracy for the in-air task is reported at 80%, which surpasses the 78% accuracy observed for the line-tracing task. Additionally, detailed metrics, including precision (referenced in Equation 3.5), recall (Equation 3.6), and the F1-score (Equation 3.7) are provided within these tables. The overall performance of the CNN approach decreases compared with the performance achieved by the KNN+DTW method. This discrepancy can be attributed to the inherent strengths of each approach regarding dataset

Table 3.2: The classification report of user classification for in-air task using robotic data.

	precision	recall	f1-score
U0	0.67	0.50	0.57
U1	0.75	1.0	0.86
U2	0.8	0.8	0.8
U3	0.5	1	0.67
U4	1	0.8	0.89
U5	1.0	1.0	1.0
U6	0.8	1.0	0.89
U7	1	0.5	0.67
U8	1	0.6	0.75
U9	0.67	1.0	0.8
accuracy			0.8
macro avg	0.82	0.82	0.79
weighted avg	0.84	0.8	0.8

Table 3.3: The classification report of user classification for line-tracing task using robotic data.

	precision	recall	f1-score
U0	0.67	0.50	0.57
U1	0.75	1	1
U2	0.62	1	0.77
U3	1	1	1
U4	1	1	1
U5	0.6	0.75	0.67
U6	0.8	1	0.89
U7	1	0.75	0.86
U8	1	0.2	0.33
U9	0.75	0.75	0.75
accuracy			0.78
macro avg	0.82	0.80	0.77
weighted avg	0.82	0.78	0.75

size, which is that KNN+DTW is good at smaller datasets, while CNNs is good at handling larger datasets.

3.3.6 Discussion

The first experiment assessed the motion-controlled robotic behaviors inherited from human operators' behaviors. In this experiment, we developed a motion-controlled robotic arm using a 7DoF Franka robot, serving as a foundational platform for our study. This experiment not only facilitated our understanding of the experimental setup and data collection processes but also prepared us for future investigations. Additionally, we designed only two task types, which provide a task prototype for our future task design. Furthermore, we deployed two algorithms aimed at distinguishing between different users, thereby establishing a solid basis for the de-

velopment of further algorithms in subsequent studies. The results of this experiment directly verified that we can classify users through motion-controlled robotic behaviors and the accuracy reaches 95%. In addition, the results on in-air and line-tracing tasks showed that performing line-tracing tasks limits human biometric identity information. Subsequently, we verified our experiments' versatility on the social robot platform.

3.4 Experimental 2: User Classification of Motion-Controlled NAO Robot

The second experiment was to classify human operators through motion-controlled social robot behaviors. It investigated whether the robotic mechanical structure, kinematic principles and the joint number would influence the performance of user classification on motion-controlled robot behaviors. Furthermore, we designed social tasks that appear in real-world human-robot social interaction scenarios and demonstrated the adaptability of our user classification approach to these social tasks.

In specific, we built a motion-controlled social robot platform. Utilizing both hands of an NAO robot, we achieved a user classification accuracy of 95.2%, yielding results comparable to those obtained with the Franka robotic arm illustrated in Section 3.3.

3.4.1 System Overview and Platform

The system overview is shown in Figure 3.9. The basic idea behind this Kinect motion-controlled NAO robot is similar to the motion-controlled Franka robot, as we described in Section 3.3.1. The human motion was captured by a vision-based tracking device (Kinect). Compared with Optitrack, Kinect is a markless and low-cost motion capture device, and it applies to the full body control of the NAO robot. Then, the captured human signals were transferred into the robotic joint angles using the inverse kinematic solver. After this, NAO executed the controlling signals and followed the human upper motion. Finally, the KNN+DTW method was implemented on both arms' end-effector trajectories of NAO. However, we used Kinect as our motion capture device. Compared with Optitrack's end-effector control, Kinect is low-cost and markless, which can be applied in more applications in different fields. Besides, the NAO robot was controlled by the human upper body, while Franka was controlled by human one arm. The Kinect tracked the human skeleton, represented by the red line on the human. The green circle was the Kinect located joint position values in Cartesian coordinates. The joint positions were transformed into commands of angular values of the robot to control the related NAO's joint with API provided by the NAOqi library.

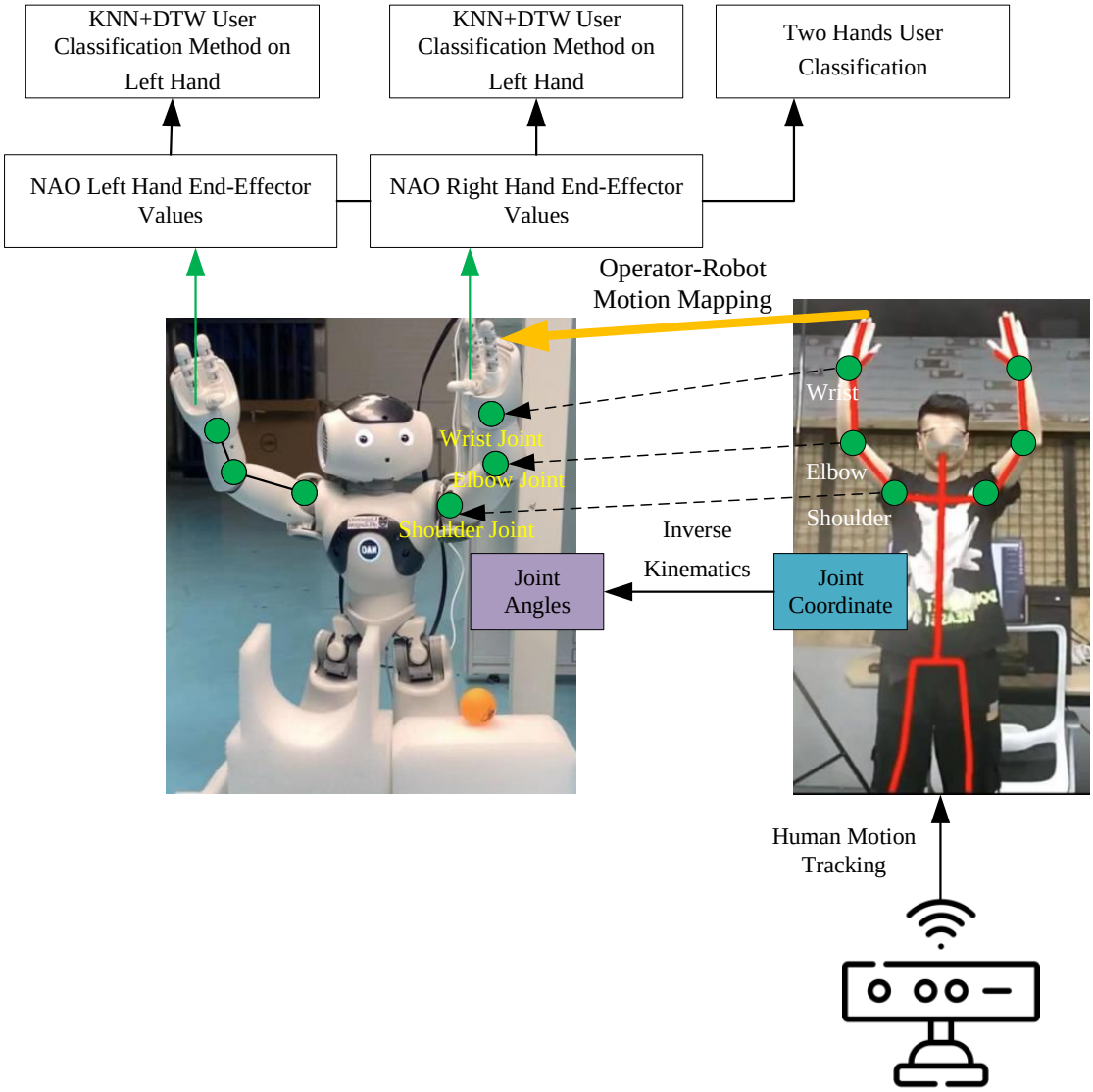


Figure 3.9: Kinect motion-controlled NAO overview.

3.4.2 User Classification Algorithm

Firstly, for NAO's upper body forward kinematics calculations, we utilized the DenavitHartenberg (DH) convention, mirroring the approach adopted for the Franka robotic arm as detailed in Section 3.3.3 and further exemplified by the transformation matrix in Equation3.11. This methodology constructed trajectories for the end-effectors of NAO's right and left arms, based on the three joint angle values for each arm. In terms of user classification, our approach was aligned with the method outlined in 3.3. Initially, the end-effector data were segmented and normalized. Subsequently, we extracted 11 features per hand, including vectors $p(t), \dot{p(t)}, ||p(t+1) - p(t)||, \ddot{p(t)}$, and $||\ddot{p(t)}||$, culminating in a total of 22 features for both hands. For user classification, we adopted the KNN+DTW algorithm, as detailed in 3.3.3, due to its demonstrated efficacy in achieving high accuracy on small datasets.

3.4.3 Experiments

In this section, we described the experiment setup and data collection in our experiments on user classification via the motion-controlled NAO platform. We had the participation of five volunteers, comprising two females and three males. All participants are college students, with ages ranging from 19 to 22 years (mean age = 21.2, standard deviation = 1.0). Before the experiment, each volunteer consented to participate and was allotted a 10-minute session to familiarize themselves with the robot's operations. The NAO robot, designed as social entities, are engineered to facilitate interaction with humans. Based on this, we selected three social tasks for the experiment, including "Waving", "Opening Arms", and "Clapping". These tasks were chosen for their relevance to social interactions and the requirement for bilateral hand coordination. Each of the five participants was tasked with executing 30 instances of each task, culminating in a total of 450 instances.

3.4.4 Results Analysis

The accuracy of classifying five users across three tasks using different robotic arms is detailed in Table 3.4. User classification was conducted separately for each hand across the three tasks. The accuracy on the left hand ranges from 93.9% to 94.4%, higher than that on the right hand ranging from 91.6% to 92.8%. Then, we combined two robot hands, by adding 22 features of two hands together. The feature number is increased and the information contained in the end-effector trajectory increases leading to the increase of the accuracy. The accuracy of combining two robot hands was higher than any of the two hands' results showing that robotic arm number can increase the user classification accuracy. A single NAO arm, with its 3 joints, offers less flexibility and results in lower accuracy compared to a single Franka arm, which has 7 degrees of freedom (DoF) and facilitates more complex trajectories. Nevertheless, the combined results

Table 3.4: Five user classification on right and left hand separately and combined for each task

	Left hand	Right hand	Average of left and right hand	Combination of left and right hand
Clap	0.9389	0.9284	0.9336	0.9516
Wave	0.9440	0.9160	0.9300	0.9470
Openarm	0.9389	0.9242	0.9315	0.9347

for two NAO arms ranged from 93.5% to 95.2%, demonstrating that they can achieve user classification accuracy comparable to that of a single Franka arm.

This outcome suggested that the limitations imposed by the fewer joints in NAO arms can be mitigated by leveraging both arms, effectively compensating for the mechanical structure and kinematic principles' constraints due to fewer robot joints. Moreover, we see that the "Clap" task achieved the highest performance on the combined hand data, which is 95.1%, while the "Openarm" task achieved the lowest performance, which is 93.5%. Through these experiments, we validated our methods on social tasks, illustrating their applicability and effectiveness in practical HRI scenarios.

3.4.5 Discussion

In this experiment, we approached user classification from a practical standpoint, focusing on a remote-controlled social robot and tasks that mimic real-world social interactions. In this experiment, we developed a motion-controlled social robot. In this work, we first showed that two hand movements cooperatively can improve user classification. Additionally, more joints such as the head joints of the NAO robot can be included to increase the user classification performance in the future. Subsequently, we investigated the control mechanisms's influence on user classification performance, focusing on the control parameters. The current model's immediate generalizability to the broader population and real-world applications is constrained by the number of participants, so future work could increase the participants' number.

3.5 Experiment 3: The User Classification under Different Controlling Parameters

The third experiment was to classify different human operators through motion-controlled robot behaviors under different controlling parameter values. From the experiment 1 and 2, we found the feasibility of our user classification method on different motion-controlled systems. In addition, we showed that the number of joints can provide more behavior information increasing the user classification performance. From the work mentioned above, I questioned whether the control capability or some control parameters that influence the robotic controlling performance

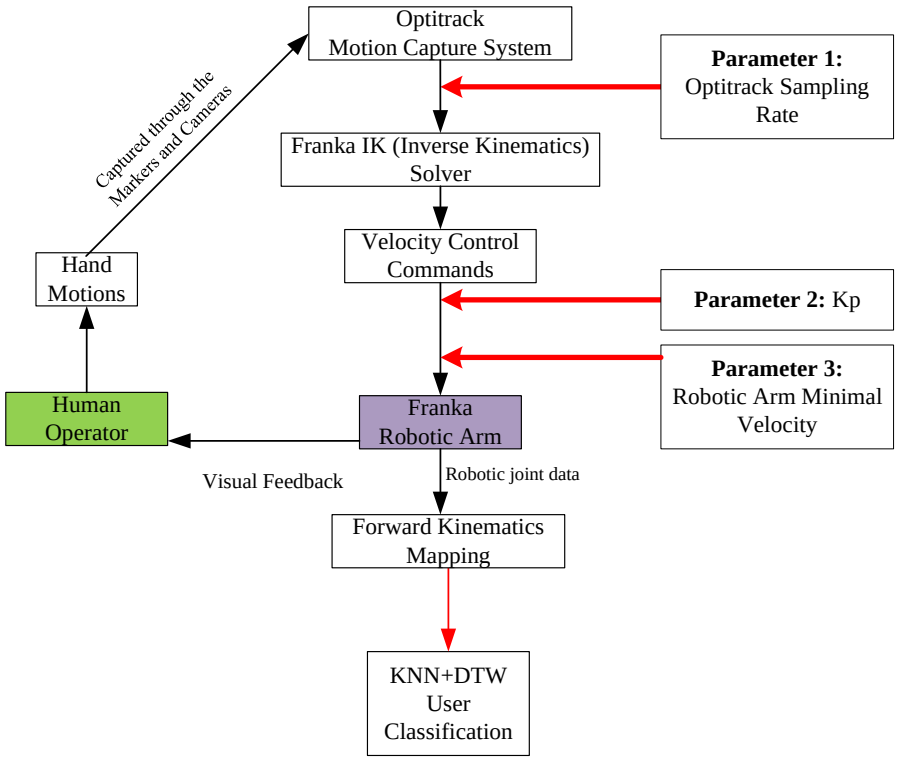


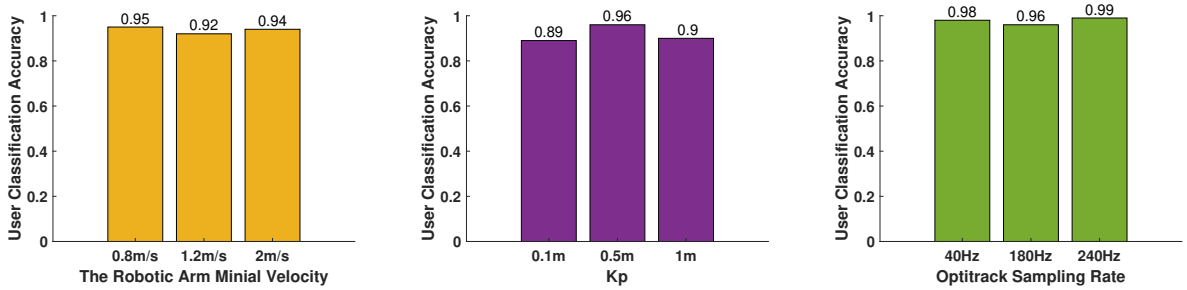
Figure 3.10: The overview of how various parameters impact user classification performance in the motion-controlled robotic system

could influence user classification performance and how they will influence them, such as three PID control parameters, including the gain K_p, K_i and K_d, as we explained in Section 3.2.1. In this experiment, we selected three main controlling parameters and investigated their influence on the user classification performance on the robotic end-effector trajectory. The three parameters are K_p, the robot arm's minimal velocity and the sampling rate of Optitrack.

3.5.1 System Overview

The system overview is shown in Figure 3.10. This study was based on the motion-controlled Franka robot platform, seen in Section 3.3.1. We adjusted the values of three parameters and implemented the KNN+DTW method on the robot end-effector data to classify different users. These three parameters have practical influence, but there is no work to study their influence on user classification performance.

K_p is the most basic component of PID control. It adds the linear error, so it gives the system a more direct and quick response. Too strong of an integral action, however, can lead to overshooting and instability. The minimal velocity for a robot arm refers to the slowest speed at which the robot can move its joints while maintaining precise control. Optitrack sampling rates decide the precision of the captured motions.



(a) User classification under three robotic minimal velocity values

(b) User classification under three K_p values

(c) User classification under three Optitrack sampling rate values

Figure 3.11: User classification results on different control parameter values

3.5.2 Experiments

The experiment was on the motion-controlled Franka arm platform. We had two volunteers, one female and one male, who were both college students. Before the experiment, each volunteer consented to participate and was allotted a 10-minute session to familiarize themselves with the robot's operations. We designed "S" as our task this time. We set 3 values for each parameter. Each participant executed this task 30 times under different control parameters' values in the air. The total number of instances was 540 ($2users \times 3parameters \times 3values \times 30times$).

3.5.3 Results

The user classification accuracy under different control parameters is shown in Figure 3.11. Figure 3.11ashows the user classification accuracy when changing the robot arm's minimal velocity from 0.8m/s, 1.2m/s to 2m/s. Figure 3.11b illustrates the user classification accuracy when Kp values were changed from 0.1m, 0.5m to 1m. The user classification accuracy is shown in Figure 3.11c when the sampling rate of Optitrack was changed from 40Hz, 180Hz to 240Hz. These values of each parameter were not most suitable for our system and did not allow our system at the best controlling performance, thus leading to the robotic more behavioral differences with operators' behaviors. Thus, we can see that the value change of these parameters leads to varying degrees of accuracy decrease. Among them, the change of sampling rate has minimal impact on the performance, while K_p leads to lower accuracy. Besides, we found that there is not a linear relationship between parameters' value and user classification accuracy.

3.5.4 Discussion

Up to this section, we found that the lower system controlling performance not only decreases the interaction efficiency but also decreases the user classification performance, thus decreasing the system's security and trustworthiness. In the future, suitable controlling parameters will be quite important to the control system design, which will help build a more secure, effective and

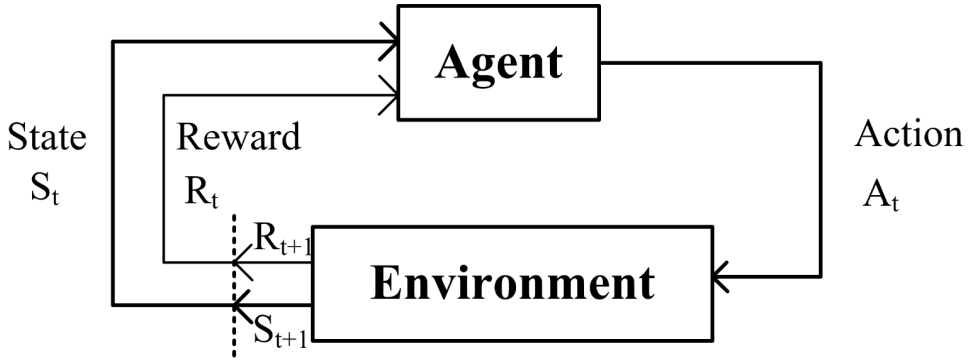


Figure 3.12: The decision processing of Reinforcement Learning

trustworthy HRI system.

3.6 Experiment 4: User Identity Protection of Motion-Controlled Robotic Arm

The fourth experiment was to implement a reinforcement learning algorithm on human-captured trajectories to dampen behavioral biometrics to protect privacy. However, in certain scenarios, human biometric information may be unnecessary to be provided. From a privacy perspective, some users may be reluctant to share their personal data with remote systems. Moreover, certain mission-critical tasks may not require the high level of personal identification that biometrics provide. For instance, during performing spine surgery, the system may not necessitate the surgeon's personal behavior data. This is particularly relevant in procedures involving young surgeons, where the priority is to offer ample practice opportunities without compromising patient safety. Based on this, the remote controlled robot should have autonomy instead of fully controlled by a remote human operator.

Shared control is a method that combines human decision and robot autonomy, which is able to assist human control intelligently. Compared with the traditional robot control, in which the robot is fully controlled by a human, there is an AI agent that processes human controlling data in the shared control system. It facilitates a nuanced collaboration that optimizes both human insights and robotic efficiency.

RL is a common algorithm used in shared control. It can learn from the environment by feeding rewards. Based on this I used Deep Deterministic Policy Gradient (DDPG) to change human hand trajectory to a standard trajectory in real time.

3.6.1 Background

Reinforcement Learning

RL is a branch of machine learning that draws inspiration from behaviorist psychology [216]. It focuses on how the intelligent agent should act within an environment to maximise a cumulative reward that has a certain measure. This area primarily concerns identifying optimal actions for the agent under uncertain conditions to achieve their goals. The learning process involves the agent interacting with its environment, performing actions, and receiving feedback from the environment by the ways of rewards and penalties. Thus, RL is distinguished by its emphasis on learning through their own actions and experiences, rather than from a labelled dataset.

RL is primarily conducted within the framework of Markov Decision Processes (MDPs). MDP is a mathematical framework used to describe decision-making problems in uncertain environments. If a problem can be formulated as a MDP, then it can be addressed using RL techniques to build models and find solutions. The core components of a LR system include the agent, environment, state, action and reward. The agent is the learner or decision-maker that interacts with the environment. The external system with which the agent interacts and which provides feedback to the agent. The state is a representation of the current situation or condition of the environment. The action is any operation or move the agent can make in the current state of the environment. The reward is the feedback from the environment in response to an action taken by the agent, indicating the value of that action. The decision flowchart is shown in Figure 3.12 RL has been applied in the field of HRI. Akalin et al. [217] provided a comprehensive review of RL in social robotics, emphasizing its effectiveness in learning optimal behaviors through environment interactions, essential for engaging with social robots.

3.6.2 System Overview

This experiment employed the motion-controlled Franka robotic arm platform, augmented with a RL AI agent that processed human trajectory data captured by Optitrack. The agent eliminated the operators' biometric identities from the data. The processed commands were then sent to the robotic arm. As a result, the users can not be classified through the robotic arm's trajectories.

3.6.3 User Identity Protection Algorithm

*Global translation:*to ensure spatial comparability among different 3D signatures, we applied a global translation to each signature. This adjustment aligned the rear-right corner of its 3D bounding box with the origin point. *DDPG algorithm:* The DDPG algorithm is a model-free, online, off-policy RL method. We set three states. They are the moving distance along the three axes of the Cartesian coordinate. The reward is the norm distance between the standard point and the state point. The Pseudo-code of the algorithm is shown in Algorithm 1. Firstly, the critic,

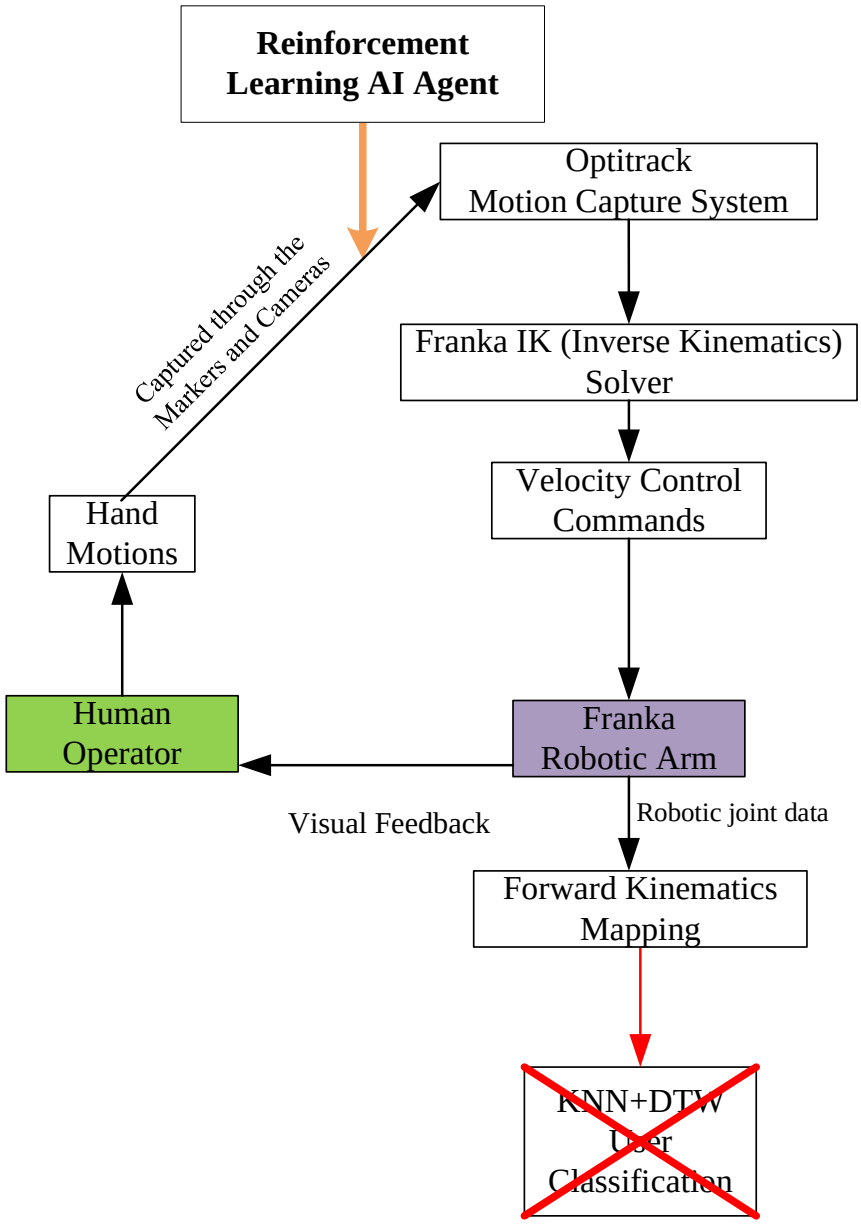


Figure 3.13: The overview of user identity protection system.

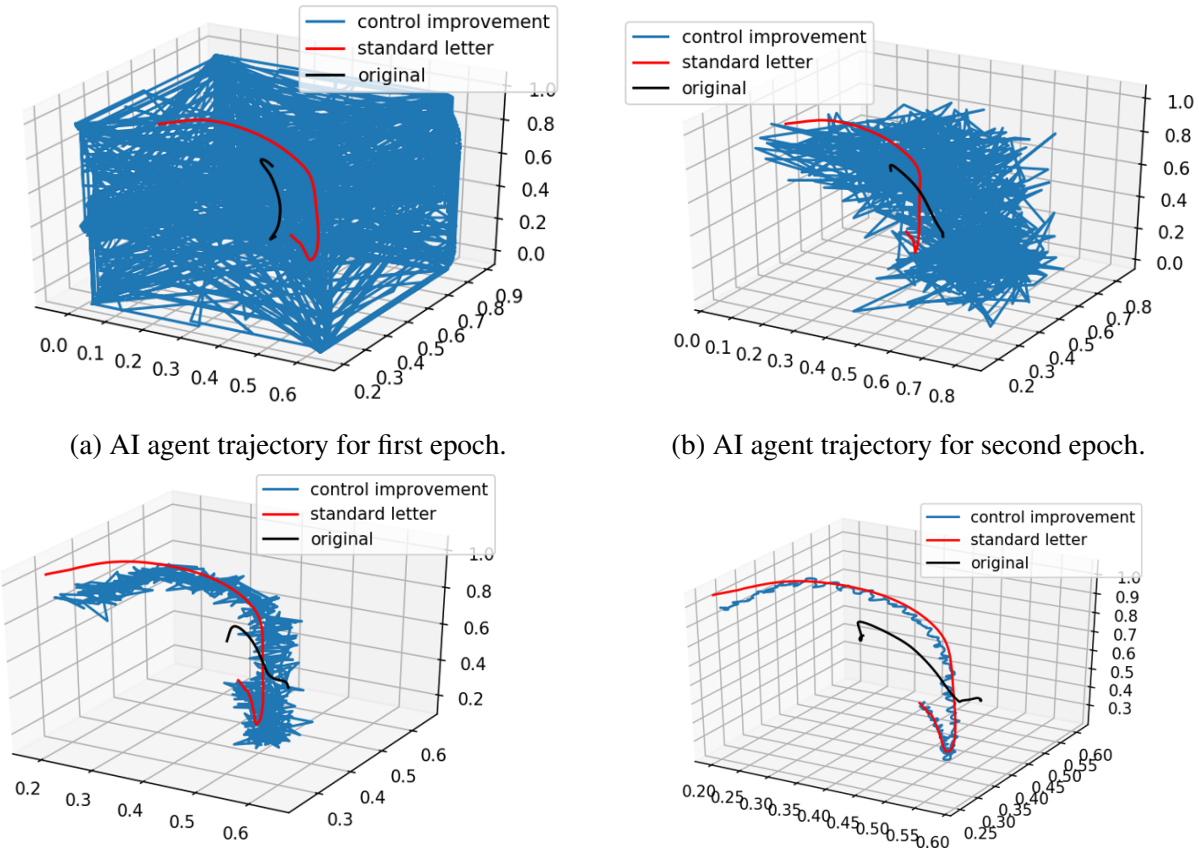
actor, target network and replay buffer are initialized. Then, for each Episode, we received the initial observation state. Action is chosen by the actor-network and is executed, after which the reward and new state are observed. Then, the transition is stored in the buffer. Then, the target Q value is calculated and combined with the prediction Q value to calculate the loss function. Finally, the critic, actor and target Networks are updated.

Algorithm 1: Pseudo-code of DDPG Model

Randomly initialize critic network $Q(s,a|\theta^Q)$ and actor $\mu(s|\theta^\mu)$ with weights θ^Q and θ^μ
Initialize target network $Q^{'}$ and $\mu^{'}$ and weights $\theta^{Q^{'}} \leftarrow \theta^Q$, $\theta^{\mu^{'}} \leftarrow \theta^\mu$
Initialize buffer R
for: episode = 1, M do **do**
  Initialize a random process N for action exploration
  Receive initial observation state s_1
  for: t=1, T **do**
    Select action $a_t = \mu(s_t|\theta^\mu) + N_t$
    Execute action a_t and observe reward γ_t and new state s_{t+1}
    Store transition (s_t, a_t, γ_t, s_{t+1}) in R
    Sample a random mini batch of N transitions from R
    Set $y_i = \gamma_i + \gamma Q^{'}(s_{i+1}, \mu^{'}(s_{i+1}|\theta^{\mu^{'}}|\theta^{Q^{'}}))$
    Update critic by minimizing the loss function: $L = \frac{1}{N}\sum_i (y_i - Q(s_i,a_i|\theta_Q))^2$
    Update the actor policy using the sampled policy gradient:
      $\nabla_{\theta^\mu} J \approx \frac{1}{N}\nabla_a Q(s,a|\theta^Q)|_{s=s_i,a=\mu(s_i)} \nabla_{\theta^\mu}\mu(s|\theta^\mu)|_{s_i}$
    Update the target networks:
    $\theta^{Q^{'}} \leftarrow \tau\theta^Q + (1-\tau)\theta^{Q^{'}}$
    $\theta^{\mu^{'}} \leftarrow \tau\theta^\mu + (1-\tau)\theta^{\mu^{'}}$
  end for
end for

3.6.4 Results Analysis

Figure 3.14 shows the RL training process. In specific, the blue line is the agent trajectory. The red line is a normalized target trajectory that does not contain user biometric identity information. The black line is the input trajectory to the AI agent, that is the user trajectory captured by Optitrack. The four figures show the AI learning process and learning results. The first epoch learning process is shown in Figure 3.14a. As Figure 3.14b shows, the AI agent (blue line) is hard to close to the target line. For the second epoch, the AI agent (blue line) found a rough trajectory as the target one (red line). Then, Figure 3.14c shows the third epoch trajectory and we can see that AI gradually found the target line and was close to the target line. Finally, the AI agent produced quite a close trajectory with the target line, shown in Figure 3.14d. During this process, AI assisted the robot on each point to move towards to standard point that has no human identity.



(a) AI agent trajectory for first epoch.

(b) AI agent trajectory for second epoch.

(c) AI agent trajectory for third epoch.

(d) AI agent trajectory for forth epoch.

Figure 3.14: Trajectory learning process of AI agent.

3.6.5 Discussions

This experiment generalized all features of human behavior to a universal level. We designed an AI agent using RL to change the human motion trajectory to a trajectory without any human biometric identity. Such an approach prioritized user privacy and increased human trust to system. Based on this work, future work could be done to protect the specific information of human identities. For instance, in the context of telesurgery, if a surgeon's distinctive hand trajectory, characterized by sharp turns, is not prefered, the system could tailor an AI assistant specific to that surgeon following user classification. This personalized assistance could significantly enhance control efficiency and operational outcomes.

3.7 Chapter3 Discussion

Chapter 3 featured a series of four experiments which explored and developed a novel category of user classification for the motion-controlled robotic arm and laid the foundation of the platform, task design, feature extraction, algorithm, experiment set-up, data collection and result analysis for developing an emotion classification system for the motion-controlled robot. User classification using the motion-controlled robotic arm end-effector trajectory was conducted in experiment 1. The accuracy of ten users on two tasks achieved 95%. However, this method was tested on one high-precision robotic arm and motion capture system, and the tasks are not specific tasks that appear in certain interaction situations. Therefore, experiment 2 built a Kinect motion-controlled NAO robot and implemented a similar algorithm for user classification across social tasks. We used two NAO arms' end-effector trajectories to perform user classification, reaching 93%. It got comparable results with one single Franka's arm, which showed the versatility of our user classification method. Whether the controlling parameters, however, will influence the user classification or not. Therefore, experiment 3 replicated experiment 1 and the study design while changing three controlling parameter values. The results showed that these unsuitable parameters not only decrease the controlling performance but also decrease the user classification performance. Therefore, after classifying the user, we can provide the most suitable controlling parameters to this user to enhance the interaction. However, in some specific scenarios, there is no need for user biometric information. Therefore, we proposed experiment 4 to train an AI agent to protect human personal information. These 4 experiments put a foundation for future human-centered HRI. More importantly, Chapter 3 lays a foundation for the next emotion classification work for motion-controlled robotic arms. They were derived from the same motion-controlled robotic trajectory data. The movements of the robot during interaction provide rich information for both identities and emotions leveraging similar features such as speed and acceleration. The generalizability of our methods to the broader population and real-world applications is limited by the number of participants, so future work could expand the participant pool. Future work could verify our method's robustness on human attackers

especially the insider attackers who observe how a user signs in space and imitate [29].

Chapter 4

Inferring Operator Emotions from a Motion-Controlled Robotic Arm

4.1 Introduction

Telerobots are a type of robotic avatar: systems operated by humans that can replicate the operator's senses and actions, allowing the operator to interact with objects in a remote location and receive relevant feedback. Utilising telerobots allows operators to overcome physical distance between themselves and a remote environment and perform actions to complete important tasks in many fields, such as healthcare and industry. Such use cases can include safety-critical and precise tasks telesurgery, nuclear waste cleaning, and remote driving, where understanding the operators' emotional state becomes crucial to avoiding dangerous outcomes. To address this, robotic systems could be imbued with emotional intelligence systems, the ability to infer and respond to human emotional states [44], in order to facilitate more effective, efficient, and engaging interactions [43] with humans operators and human bystanders. For example in remote driving scenarios, a teledriver could be warned to take safety measures when significant fatigue or stress is detected. Further, understanding the user's emotional state could allow intelligent re-

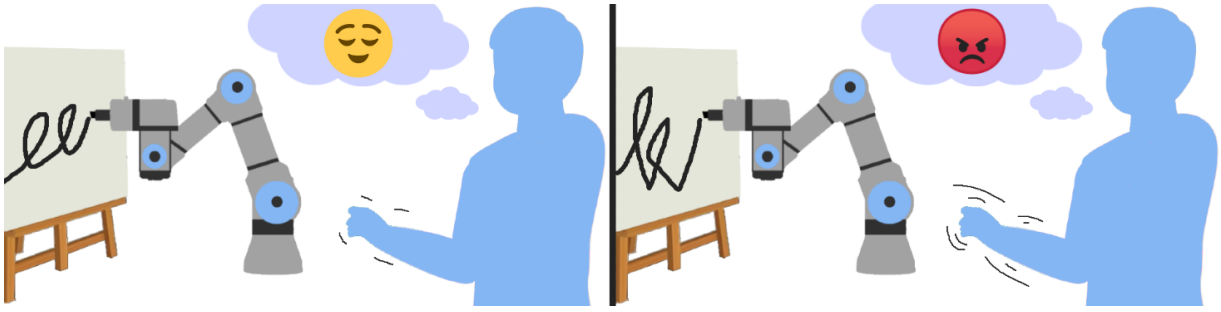


Figure 4.1: The emotions of a human impacting the trajectory of the motion-controlled robot arm they operate. We investigate how the robot's movement can be used to infer these operator emotions.

mote control algorithms to automatically adapt to potential emotional actions, improving control efficiency and mitigating negative outcomes. For example, such a system could assist surgeons with crucial emotion regulation [218]. While surgeons undergo self-assessment procedures, their efficacy is impacted by a stressful environment and surgeon's level of experience [219]. When using a telerobot to conduct surgery under intense or suppressed emotions, imprecise or exaggerated movements may result. Our approach could complement self-assessment as a safeguard, detecting in-the-moment emotional change, which could in turn be used to normalise emotional movements. Finally, this work could lay the foundation for understanding and facilitating more emotional and expressive encounters between co-located humans and telerobotic avatars in work and social contexts.

Existing approaches that allow robots to classify human emotions rely on analyzing individual human status data, including the physiological and behavioural signals [220]. Prominent physiological affective measures based on electrical signals include electrocardiography (ECG), electroencephalography (EEG) and electromyography (EMG), while behavioural signals include facial expressions, bodily movement, and speech signals [60]. These methods can, however, be unsuitable for telerobotic avatar operation. Movement by the operator to control the robot can interfere with data collection, as wearable sensors may struggle to maintain a stable position to capture reliable physiological signals [60] and electrical signals are sensitive to movement artifacts. Furthermore, cameras may fail to capture facial expressions due to the visual occlusion [221] and speech recordings may be unavailable or indistinct in work settings [43]. Additionally, these methods require users or their workplaces to be burdened with additional devices [60, 222] and the data captured, such the user's physiological signals and facial expressions, can contain highly sensitive information, raising privacy issues.

To address these concerns we propose an alternate approach, inferring the human operator's emotions by studying the movements of the robotic avatar. Bodily movements, such as the movement of hands and head, are emotional cues that can communicate emotional status [223] and work by Huang et al. [224] has demonstrated that a motion-controlled robot can inherit the human operator's motion behaviours. Building upon this prior knowledge, we verified this finding and leveraged it to present a first-of-its-kind system that automatically detects a human operator's emotions based solely on the movements of the robotic avatar they are controlling. We developed both a physical motion-controlled robotic avatar platform and learning-based emotion recognition algorithms to analyze the joint and end-effector readings of the avatar's non-stylized motions.

Two task types representing different remote control scenarios were used: 1) mid-air gestures to represent general industrial telerobot scenarios [225] and gesture tasks representing social scenarios [44, 138], 2) a line-tracing task to represent safety-critical scenarios [215]. The participants listened to affective audio files, such as symphony, noise and comedy, to stimulate different emotions [104, 226], while controlling the robotic arm to perform tasks and data

was collected from both the robotic avatar and an ECG device fitted to the participant. We developed a Dynamic Time Warping (DTW)-based algorithm and a Convolutional Neural Networks (CNN)-based algorithm to recognize the user's emotions, while the training model can be subject-independent or subject-dependent. We further derived unique features from the robotic arm's movement to capture the human user's emotion inherited by the robotic avatar. The emotion recognition accuracy among five emotional states of our method reaches to 83.3%, while the accuracy of the ECG-based classification was 54.6%. We finish by discussing the implications of our method on telerobotic applications both present and future. The contributions of our work are summarized as follows:

- We demonstrate for the first time that a motion-controlled robotic avatar can inherit human affective states, achieving an emotion recognition accuracy of 83.3% across five emotions based on the robot's movements.
- We developed two alternative emotion recognition algorithms based on DTW and CNN, respectively, and further developed unique emotional features from the robotic avatar's end-effector motions and the robot joints' spatial and temporal features.
- We demonstrate through direct comparison that our approach is more effective and appropriate for motion-based telerobotic applications than traditional ECG-based methods.
- We discuss the implications of our approach and findings on current and future Human Robot Interaction (HRI) applications.

4.2 Background

4.2.1 Modelling Emotion

Prior works [61, 74] classified the emotion models into four categories including dimensional, discrete, meaning-oriented, and componential emotion models. Dimensional emotion models include uni-dimensional and multidimensional models. Russell popularized a multidimensional model [61, 227], proposing a two-dimensional circumplex emotion model in which the x-axis represents valance and the y-axis represents arousal. Valance represents the hedonism of emotion, whether it is positive/good or negative/bad, while arousal represents how alerting/attention-grabbing/exciting the emotion is. For example, joy is a high-arousal and positive emotion, while sadness is a low-arousal negative emotion. Ekman [61, 75] popularized the discrete model and proposed basic emotions. Meaning-oriented models use lexical meaning and social constructivism to represent emotion, while componential models represent emotions by elicitation of emotional responses. There are two reasons why we chose to use Russell's model in this work [61, 227]. First, the majority of research on emotion recognition uses the circumplex

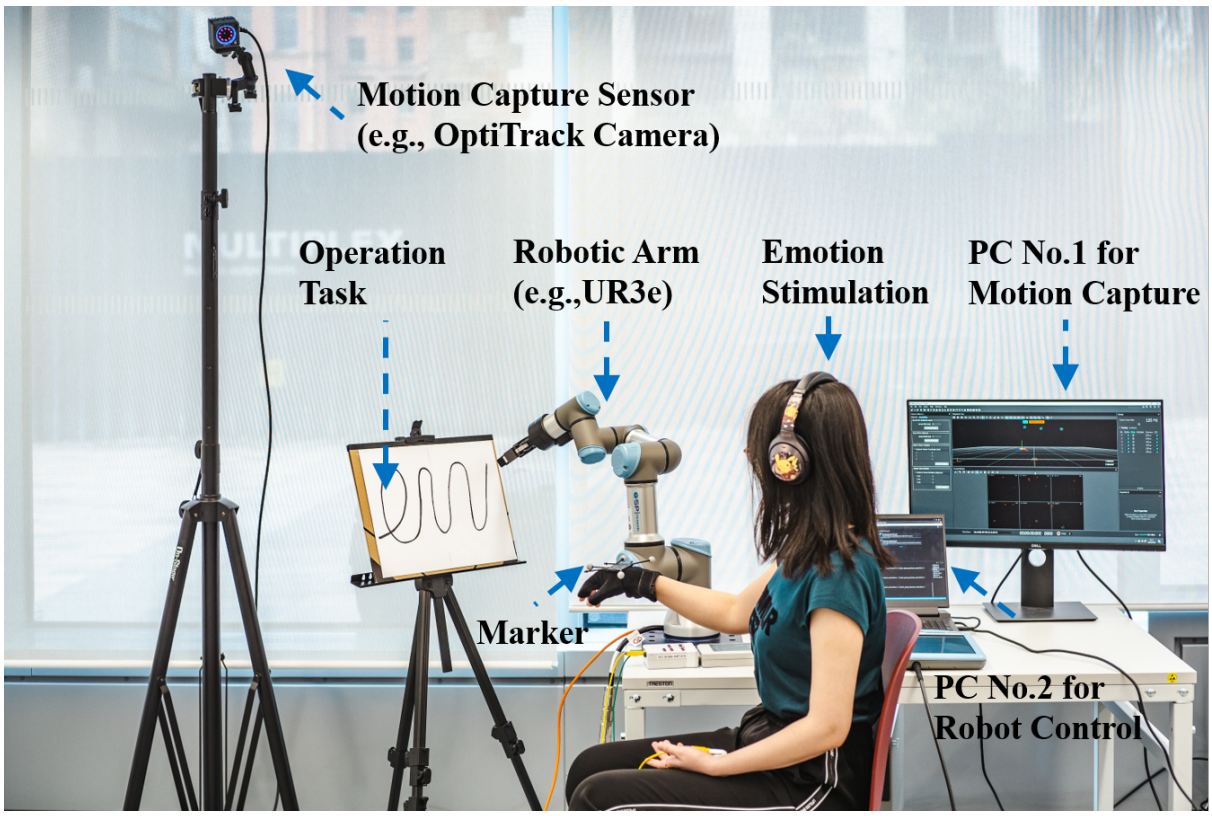


Figure 4.2: Our robot platform with motion-based emotion transmission.

model, allowing for external validity and comparison [60, 104, 228]. Secondly, the circumplex model allows for a more holistic and less discrete spectrum of emotions, as opposed to selecting a limited set of emotions. In this thesis we induced and observed emotions from each of quadrant of the circumplex model (Joy, Sadness, Anger, Pleasure) with the addition of a neutral emotional state between them, an established approach used across many prior works [44, 104, 228–232]. We visulised the emotion model shown in Figure 4.5

4.2.2 Emotion Recognition during Human-Computer and Human-Robot Interaction

Emotional intelligence is the ability to recognize and generate emotions [44]. Endowing computers and robots with emotional intelligence could enable more intuitive, efficient, and collaborative human-computer and human-robot interaction [43, 44, 233, 234]. By enabling an intelligent computer or robotic agent to infer human emotions, the agent could give corresponding feedback, such as activating alarms and generating expressive behaviors [234]. The application of emotion recognition includes health monitoring, user experience assessment, intelligent assistance, social interaction [220], education, surgery [235], and robot rehabilitation. The data utilized for emotion recognition can be categorized into two types: human individual status data and interaction information left on computers or mobile devices [220]. Below, we summarize

the existing works which implement emotion recognition using these two types of data.

4.2.3 Emotion Recognition Using Human Individual Status Data

The human individual status data can be further divided into two categories: physiological signals [60] and behavioural signals [223].

Physiological Signals

Emotion recognition using physiological signals is a hot topic [88]. Physiological signals include electroencephalography (EEG), electrocardiography (ECG), HRV, galvanic skin Response (GSR), respiration rate analysis (RR), skin temperature Measurements (SKT), electromyogram (EMG), and electrooculography (EOG) [60]. Among these, EEG and ECG are most frequently used for emotion recognition [61]. EEG records the electrical activity of the brain by placing electrodes on the head, using 8, 16, or 32 electrode pairs in most cases [60]. ECG detects the electrical activity of the heart by attaching three electrodes around the body [60], while Zhao et al. [228] has also proposed using a wireless device to capture ECG signals. There are limitations to these techniques, however. Human motions produce motion artifacts and interfere with inferring from electrical signals for both EEG [159] and ECG [101]. Based on this, it is advised not to collect EEG and ECG data when participants are moving, rather authors like Dzedzickis et al. [60] advise that EEG and ECG should be administered when the subject is calm and stays stable. Similarly, EMG and EOG, which have been used to detect electrical signals of muscle cells and eye movements respectively, can be influenced by motion artifacts. SKT is limited by the latency between emotion generation and skin response [60]. These limitations provide motivation for an emotion inference system usable in the movement-based scenario of telerobotic operation. To compare the suitability of our novel approach with an established technique, we utilised ECG in this study.

Behavioral Signals

The behavioural signals can be divided into two types, verbal signals [236] and non-verbal signals [223], where non-verbal signals further include facial expressions and bodily movements. Voice signals and facial expressions require additional devices and abundant computing resources to process in real-time and are hard to capture while humans are moving. Recognizing emotions from gestures and bodily movements remains an under-explored and underestimated topic [140, 223]. Emotion-related features that can be extracted from kinematic features of bodily movement (e.g., head, arm, upper body or the whole body) and expressive features [237]. Kinematic features include velocity, acceleration, and jerk of trajectory [44, 138, 142, 235, 237], while expressive features include spatial extent, energy, symmetry

and leaning of the head [237, 238]. Speed is related to how energetic the movement is, acceleration indicates muscle tension, and jerk represents the force [239]. Prior emotion recognition research has used average hand speed, acceleration and jerk [102, 142]. 14 joint velocities, acceleration, time duration, as well as the mean and standard deviation values of velocity and acceleration, were also used [61, 235]. Pollick et al. [138] found correlations between the kinematics features of the arm and emotion model space, and this finding is across many prior works and the years [44, 240]. In particular, correlations were found between higher arousal levels and several other factors: shorter duration, greater magnitudes of velocity, acceleration and jerk the movements have. Another correlation was found between positive, higher valence emotions and kinematic features with smaller magnitudes and longer levels of duration. There are, however, privacy concerns when using human facial or movement data directly to infer emotions, as sophisticated camera setups may be required and detailed live video data sent over networks for remote processing. By instead inferring emotion from a robotic avatar, one could bypass this invasive step.

4.2.4 Emotion Recognition via Interaction with Desktop and Mobile Device Interfaces

Human behavior when interacting with computer interfaces can also be used to infer emotions. For example, the user's typing content in X (Twitter) [241], phone calls, browsing websites, location information, and the frequency of app usage in mobile phone [220] can be used to analyse and recognize different emotions. Similarly, Andreasson et al. [50] used human touch information on a robot to classify emotions and other work has demonstrated that mouse movement and touchscreen dynamics can also reveal the user's emotional state [242, 243]. Besides, mouse and keystroke dynamic information can be used for emotion recognition as well [242]. However, these methods all require a long period of data collection and cannot recognize emotions in real time, limiting their usefulness to safety-critical applications.

4.2.5 Robot Emotion Expression

Prior work has explored autonomous robotic emotion expression across different form factors. For example, Ghafurian et al. varied the movements of body parts such as the tail, ears, eyes and head of the animal-like robot Miro [244] to express emotion. Saerbeck et al. [56] explored how a vacuum robot's movement can convey emotion while others have explored adjusting the motion parameters of humanoid robots, including acceleration, velocity, and curvature [245, 246]. They showed that robots have abilities to express emotions through their motions and there exist relations between motion parameters and emotions. Empowering robots with emotional intelligence could endow robots with the ability to not only recognize emotions but express emotions. The robot's ability to express emotions can greatly influence the resulting social

interaction [43]. Following emotional inference, robots could adjust their emotional display to show empathy or positively influence the emotions of the user. For example, when the user is sad, a robot could attempt to induce happy emotions to comfort them. While prior work has explored the emotional expression of social robots, we present novel findings on how emotions manifest in the movement of robotic arms used in industrial or medical settings, paving the way for more affective interactions between humans and operated or autonomous robots in current and future human-robot workplaces.

4.3 System Overview

In this section, we introduce a motion-controlled robotic avatar platform and present an initial feasibility study to demonstrate that a robotic avatar can inherit the operator's emotions. Additionally, we introduce the architecture of the proposed emotion classification approach.

4.3.1 Motion-controlled Robotic Avatar Platform

We built a motion-controlled robotic avatar platform, where a human interactively controls a robot arm, as shown in Figure 4.2. An OptiTrack system [209] was built using six cameras arranged in a circle to capture the operator's hand motion trajectory via a glove attached with four markers. A personal computer (PC1) calculated the 3D hand coordinates and sent the pose data to a second PC (PC2). PC2 calculated the position velocity and angular velocity via the received 3D hand trajectory data from OptiTrack. PC2 was also installed with the robot operating system (ROS) [247] which controlled the Universal Robot UR3e [248] in real-time with the help of MoveIt, a 3D motion visualisation and control software platform [249]. During operation the human receives the visual feedback of the robot's current position and adjusts their hand motions for interactive control, leveraging hand-eye coordination.

4.3.2 System Architecture

The basic aim of our system is to use the robotic avatar's motions to infer the operator's emotions during interactive control. The architecture of the proposed robotic avatar emotion classification approach is illustrated in Figure 4.4. An operator controls the robot to perform motion tasks while in different emotional states. The operator's hand motions are first captured by a motion capture system and further calculated to control command sequences that are sent to the robot to enable the robot to execute tasks in real time. Meanwhile, the operator observes the robotic avatar's motions and adjusts his/her own motions to perform interactive control, using hand-eye coordination. Two emotion classification methods are deployed on the robotic avatar. When the robotic avatar executes the motion tasks, the values of the robot's joints and the end-effector data (the endpoint movement of the robotic arm) are input to our classifiers.

Segmentation and Calibration is first applied to acquire the instances of the performed motion tasks. In order to observe how the operator's emotion influences in the robotic avatar's motions, we developed two emotion classification algorithms, a DTW-based algorithm, and a CNN-based algorithm. In the DTW-based algorithm, segmented end-effector trajectory is used, after which *Emotion Related Feature Extraction* derives unique motion features to capture the operator's emotion information. The derived features are then normalized and analyzed by DTW to infer the operator's emotion. In the CNN-based algorithm, the segmented time series of all the robot's joint rotation angles are analyzed and mapped into polar coordinates by *Joint Trajectory Mapping* to generate colour gradient polar plots, with different colours to present different joints and a light to dark gradient to present time (see Figure 4.8). This approach presents both spatial and temporal features of a task instance as a 2D image. These colour gradient plots are evenly split into training and testing datasets and fed into a CNN-based model for emotion classification. Finally, based on the classification result, we can infer the operator's emotions. At this stage, a real-world system could decide whether to continue or abort the operation according to the inferred emotions and the importance of the current task.

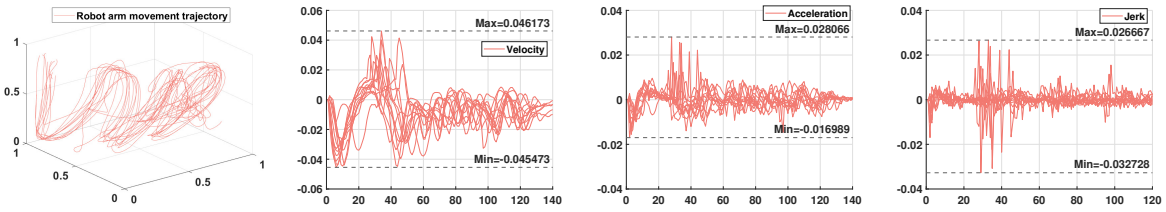
4.4 Feasibility Studies

In this section, we did two feasibility studies on two different motion-controlled robotic arm platforms. One is conducted on the Franka robotic arm and the other one is conducted on the UR3e robotic arm. These two studies verified that the motion-controlled robotic arm can inherit human operators' behaviors and showed that the robot can inherit the operator's emotions. Besides, we showed that 20 emotion-related features can distinguish two emotions and the three emotion classifications reached to 89%. The feasibility studies provided a solid foundation for our emotion classification study.

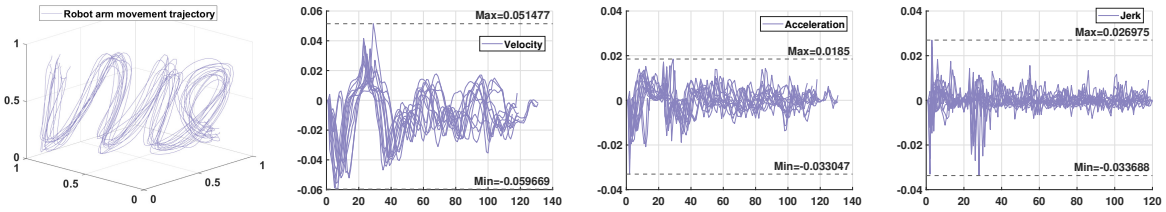
4.4.1 Inheriting human behaviours of the motion-controlled robotic arm

Emotional state impacts our behaviours to varying degrees, and emotional changes can be detected through human motion behaviors [250, 251]. Can these emotional changes be inherited by the robotic avatar? Huang *et al*. [224] showed that a robot could inherit the human operator's behaviours and human behavioural biometrics are embedded in the robot. We verified that the robot can inherit behaviors in Section 4.4.1, however, it was unclear at this stage if the affective state is also inherited.

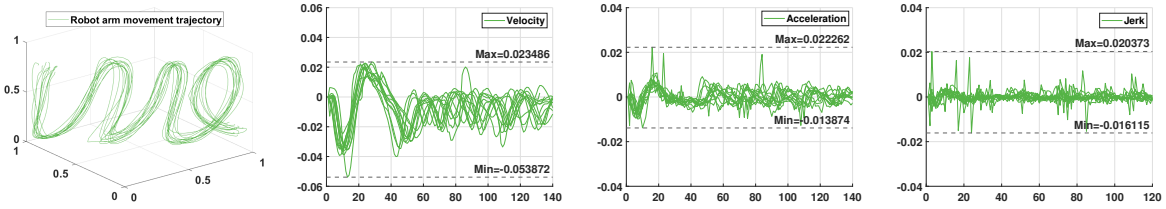
This pre-study investigated whether the robotic avatar can inherit the human operator's emotions such that it manifests in distinct differences in motion. One participant, a twenty-five-year-old female volunteer, was asked to draw "Lw" repeatedly according to a reference trajectory on a whiteboard while controlling the robot. At the same time, the participant was asked to listen to audio files to stimulate emotions during the experiment. The audio files were picked up by



(a) Joy.



(b) Annoyance.



(c) Neutral.

Figure 4.3: Robot's 3D movement trajectory, velocity, acceleration and jerk plot for joy, annoyance and neutral emotions.

the participant in advance. This audio induction method is a well-established approach including the use of music [104] and sound [226]. For the pre-study three emotional states from the circumplex model were used: joy, annoyance and a neutral/baseline emotion. The participant repeated the drawing task 12 times under each emotion, respectively. Figure 4.3a, Figure 4.3b, and Figure 4.3c show the robotic avatar movement trajectory, velocity, acceleration, and jerk plots under three different emotion conditions. The illustrated figures show the difference in the robot's motion information between different emotions.

We observed that, when the operator was emotionally neutral, the robotic avatar's trajectories were less frenetic or dramatic, while the joyful and annoyed trajectories featured more sudden shifts in motion, as well as greater peaks and troughs. More specifically, the widest range of trajectory velocities occurred during annoyed emotional states, followed by joyful states, with neutral states having the smallest range. Trajectory acceleration was more consistent during neutral states than joyful and had fewer fluctuations than during annoyed states. Joyful and annoyed trajectories were more disordered and erratic than neutral and jerks were more common. This indicates the operator was either less concentrated on their motions when influenced by these stronger emotions [252, 253], or that these high arousal emotional states [61, 227] caused with the operator to subconsciously performing stronger, more active and more erratic gestures, an effect identified by Glowinski *et al.* [237], Pollick *et al.* [138] and Wallbott*et al.* [238]. These effects can also be observed in mouse movement and touchscreen dynamics [242, 243]. Initial findings from this pre-study suggested that when operators express higher arousal emotional states, their motions become more energetic and less stable, further suggesting that emotional state can influence a robotic avatar's trajectories.

4.5 Emotion Classification Algorithm Design

In this section, we introduce the adopted emotion model and give detailed descriptions of methods for feature selection and emotion classification.

4.5.1 Dimensional model of emotions

Our work utilised the two-dimensional emotion model established by Russell *et al.* [254], a common approach for researchers [60, 104, 228, 255] (see Section 4.2.1). The x-axis represents the valence, and the y-axis represents the arousal of the emotion. We used five basic emotions defined in each of the four quadrants of the model respectively: joy, pleasure, sadness, annoyance and a central neutral emotion [44, 104, 228]. Figure 4.5 is 2D coordination picture, where the x axes is valence and the y axes is arousal. We locate the four basic emotions to represent in each quadrant respectively and the neural emotion in the center using the orange cross symbol. Joy is categorized in the positive valence and high arousal quadrant, while pleasure is categorized in the positive valence and low arousal quadrant. Annoyance is categorized in the negative

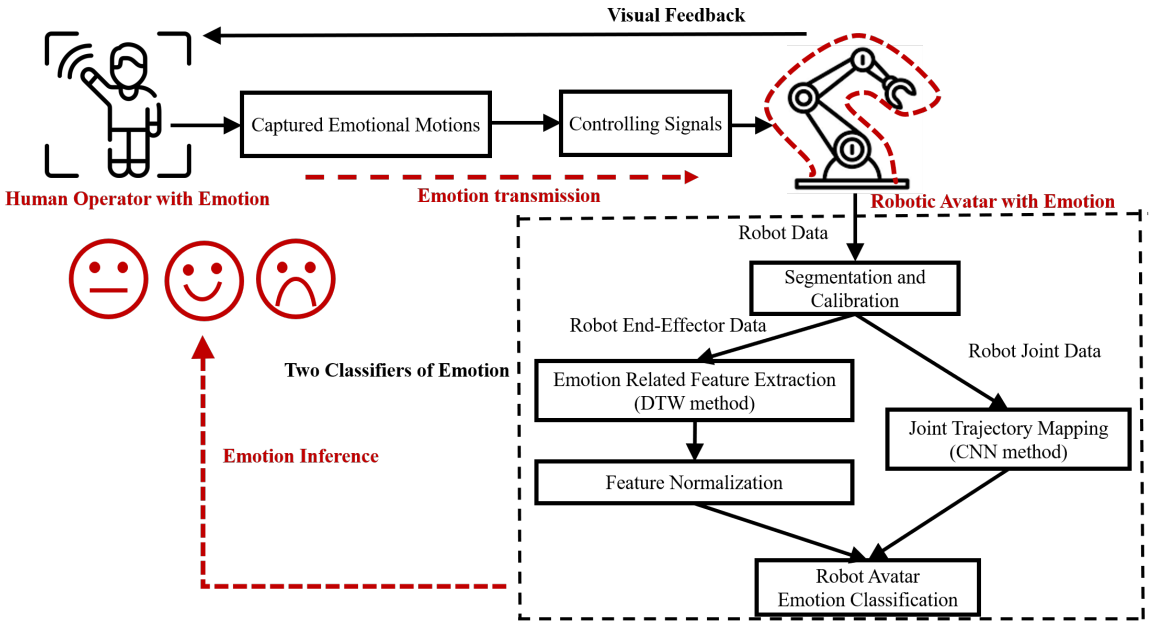


Figure 4.4: The architecture of the proposed human emotion inference through the robotic avatar. The red lines represent the transmission process of emotions. In specific, the operator performs emotional hand motions, which are executed by the robotic avatar in real time. During motion transmission, the emotional contents in motions are also transmitted. Then, these emotional contents are classified and the operator's emotion is inferred.

valence and high arousal quadrant, while sadness is categorized in the negative valence and low arousal quadrant. Our method analyses these five emotions, as shown in Figure 4.5.

We collected five basic emotions that were used across many prior works and the years in the field of affective computing as explained in Section 4.2.1. In the future, we can analyse non-basic emotions, such as confusion, frustration, boredom and engagement, which may occur when participants interact with robotic interfaces in the real world [256,257]. Future work could also build on our work to test how well it can apply to real HRI contexts, such as medicine and education, as in different contexts the requirements and benchmarks for emotion recognition may change. For example, in telesurgery emotion recognition should focus on the intense emotional status, while in e-learning emotion recognition may focus on more positive and negative emotion detection.

4.5.2 Data Segmentation, Normalization, and Calibration

The robot's positional data was first segmented into instances based on the end-effector's trajectory. Specifically, we set a threshold to the velocity of the end-effector's trajectory and use this threshold to determine each instance's starting and ending positions. Each repetition of one completed task is regarded as one instance. The x, y, and z axes of segmented instances are then normalized into a $1 \times 1 \times 1$ bounding box to make them comparable to each other. In addition, the starting position of the trajectories was aligned with the origin of the UR3e to make the

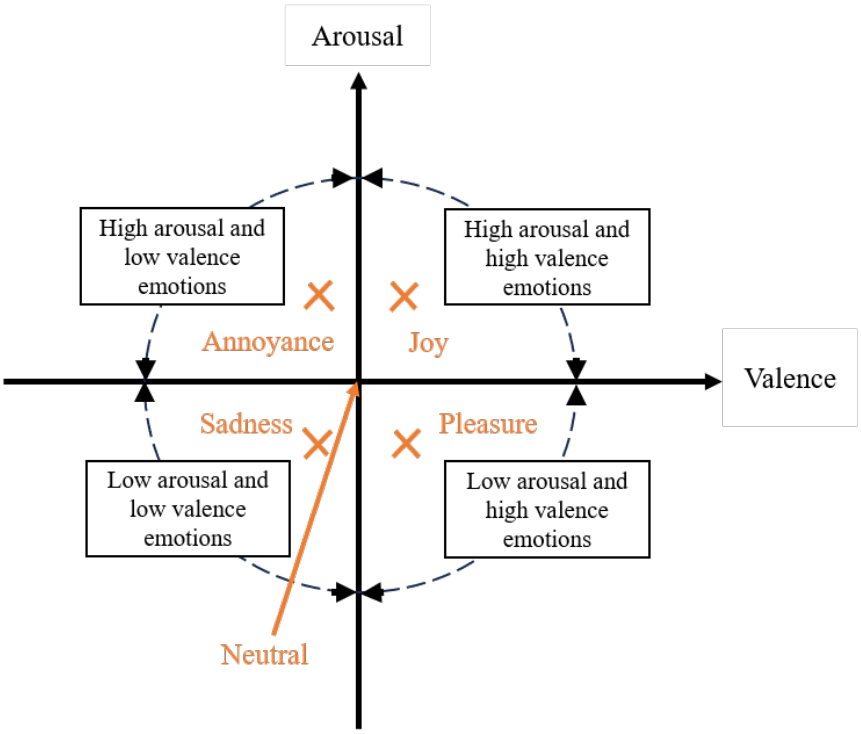


Figure 4.5: The four emotional states we induced in this work, Joy, Pleasure, Annoyance and Sadness mapped to each of the four quadrants of Russel's circumplex model of emotion, with neutral at the origin.

Table 4.1: Emotion features of robotic end-effector trajectory for DTW

Kinematic Features	3D position	p(t) = (x(t), y(t), z(t))
	3D velocity	$\dot{p(t)} = \frac{dx}{dt} + \frac{dy}{dt} + \frac{dz}{dt}$
	3D position difference	‖ p(t+1)-p(t) ‖
	3D accerlation	$\ddot{p(t)} = \frac{d^2x}{dt^2} + \frac{d^2y}{dt^2} + \frac{d^2z}{dt^2}$
	3D acceleration norm	$\| \ddot{p(t)} \|$
Expressive Features	Time range	*t(N)-t(1)*
	Energy	H=-$\sum (P(pos(t)) * log2(pos(t)))$
	Spatial extent	$\sqrt{Range(pos_x(t))^2 + Range(pos_y(t))^2 + Range(pos_z(t))^2}$
	Jerkness	$\dddot{p(t)} = \frac{d^3x}{dt^3} + \frac{d^3y}{dt^3} + \frac{d^3z}{dt^3}$
	Curvature	$\kappa(t) = \frac{\sqrt{c_{zy}^2(t) + c_{xz}^2(t) + c_{yx}^2(t)}}{(\dot{x}^2(t) + \dot{y}^2(t) + \dot{z}^2(t))^{3/2}}; c_{zy}^2(t) = \ddot{z(t)} \times \dot{y(t)} - \ddot{y(t)} \times \dot{z(t)}$
	Slop angle	$\beta_{xy}(t) = \arctan \frac{\dot{y(t)}}{\dot{x(t)}}$ $\beta_{zx}(t) = \arctan \frac{\dot{x(t)}}{\dot{z(t)}}$

instances spatially comparable.

4.5.3 Robotic Avatar Emotion Classification by Using DTW

We developed a DTW-based algorithm to classify the robotic avatar-inherited human emotions, which utilizes the positional data of the robot end-effector within the segmented instances.

Emotional Related Feature Extraction

In order to capture the operator's emotions manifesting in the robot motions, 20 emotion-related features were extracted from the robot end-effector time-sequenced data. They are donated as $p(t), \dot{p(t)}, ||p(t+1) - p(t)||, \ddot{p(t)}, ||\ddot{p(t)}||, \dddot{p(t)}, \beta(t), \kappa(t)$, where the t means the time sampling index, $t = 1, ..., N$. As illustrated in the following table 4.1, we provided mathematical formulas of emotion features analyzed in our study.

These features can be categorized into two types. The first type is kinematic features, including the robot's 3D trajectories, 3D velocity, 3D acceleration, 3D jerkiness, and position difference. The second type is expressive features, including slope angle, curvature, energy, spatial extent, and time range. We chose to examine motion features to further establish whether distinct differences in trajectory can be observed between a wider range of emotional states, following promising early results from our pre-study (see Figure 4.3). It was natural to investigate the expressive features, including energy, spatial extent and time duration, as these have been shown in prior work to convey emotional information in other contexts [237]. We calculated the energy of trajectories by calculating the entropy of signals and the spatial extent of each task

instance by calculating the size of each trajectory. Higher energy motion relates to high arousal emotion, while lower energy relates to low arousal energy [61, 237, 238]. The motions' use of space indicates valence of emotions [61, 237, 238]. The time range is a key factor in judging the human emotion's arousal level [258], so we calculated the time of performing each trajectory. In addition, we extracted jerkiness, slope angle, and the curvature of trajectories to represent the motion smoothness, as Glowinski *et al.* [237] showed that smoother movement correlated with emotional expressions.

The Emotional Classification Capability of these Features:

After conducting two rounds of analysis, the first being a pre-study 4.4.1 where we presented figures illustrating the direct robotic kinematic features to examine the relationship between emotions and motion trajectories. The second focusing on the theoretical foundation underlying the connection between specific features and specific emotional states, as well as the broader relationship between motions and behaviors, where we proposed 20 emotion-related features. Based on this, we implemented two methods to evaluate the selected emotion features, including one DTW boxplot method and one PCA visualization method.

Boxplot Method:

We used the DTW algorithm to evaluate the discriminative ability of these features. Firstly, we selected two emotions with substantial differences: neutral and annoyed emotions. Then, we split the 60% of date sets into training sets and 40% into test sets. We used DTW to iteratively compare the distances between two instances of the training dataset and computed the values. After this comparison, we located the instance that calculated the smallest distance and regarded it as the template instance. In the test datasets, we compared each instance with the neutral emotion template and the annoyed emotion template respectively. We applied this method to all participants and they showed similar results in emotions. Figure 4.6 shows an example of one subject's feature comparison results in two emotions. In specific, there are 40 boxes in total in this boxplot and for each emotion feature, there are two boxes, where the outline one and the black-filled one represent the distance between the annoyed emotion template and the neutral emotion template, respectively. For each box, there are a median value and the box width. The figure shows that 20 features show significant distance discrepancy in regard to these two emotions. It shows that the DTW distance difference is quite large between two emotions for time range and energy feature. Besides, velocity, acceleration and jerk in three dimensions show larger distance differences, compared with the other features.

PCA Visualization Analysis:

Principal Component Analysis (PCA) is a statistical technique used to reveal the underlying structure of data and preserve the maximal information by reducing its dimensionality, while preserving as much of the original data's variability as possible [259]. Simply put, PCA seeks to identify the most important features in the data, transforming them into a new, smaller set of features called principal components. This method is commonly used in data preprocessing,

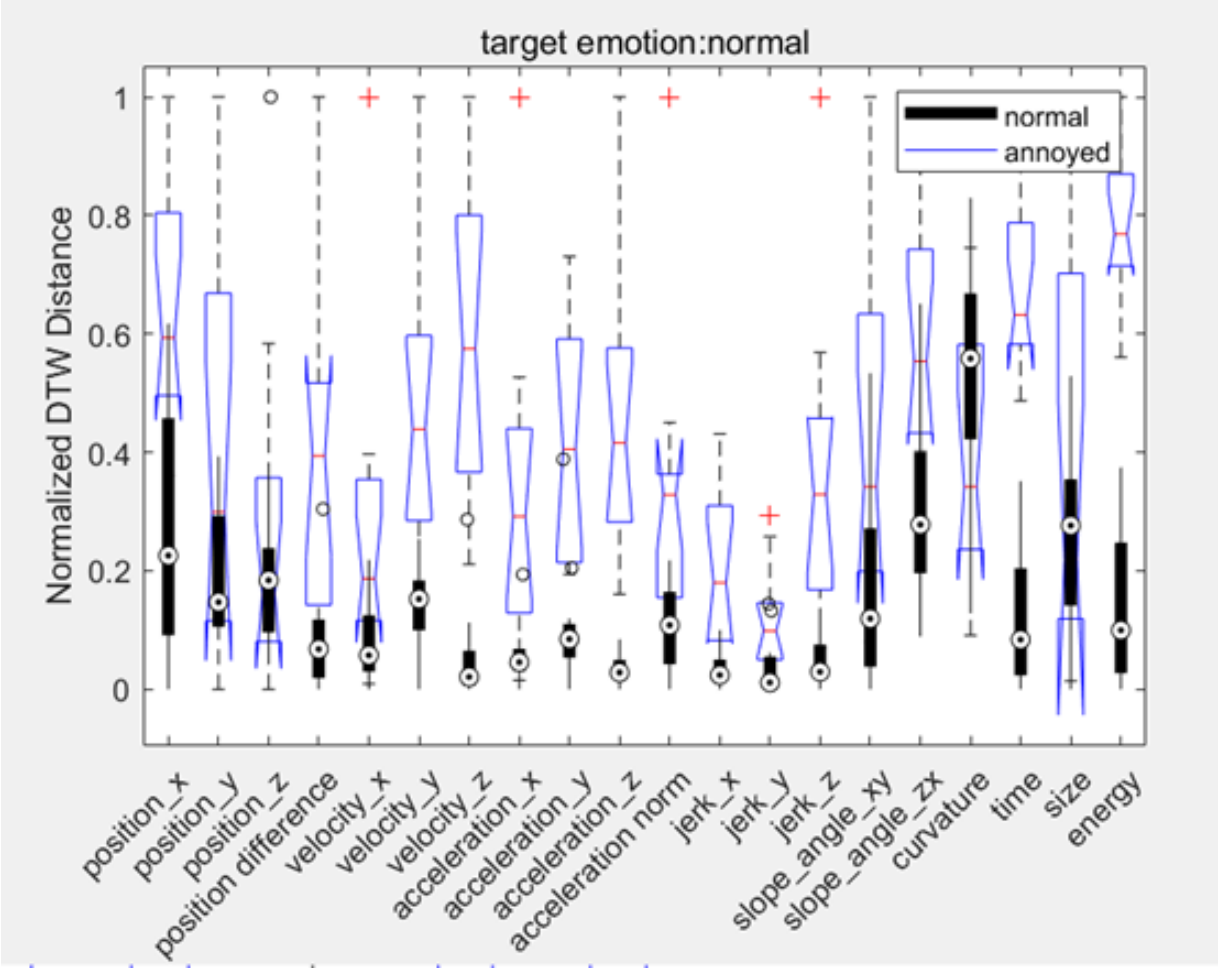


Figure 4.6: The normalized DTW distance between neutral and annoyed emotions.

data compression, and visualization, especially when dealing with high-dimensional datasets. The steps of downsampling PCA can be summarized as follows:

Algorithm 2: Downsampling PCA

Paramerers: Number of dimension K.
Input: Dataset as matrix X $\in R_{N\times M}$
Output: Eigenvectors W $\in R_{M\times K}$
$X^{'}$ = RandomColumnSampling(X)
U_1BW_1^T = Householder($X^{'}$)
U_2CW_2^T = Diagonalization(B)
W = W_1W_2

It was applied to these features to validate whether they have the ability to distinguish different emotions. Specifically, we calculated the mean, variance and standard deviation values for each feature sequence, resulting in 39 static feature values in total for each instance. Then we reduced this 139 vector to 12 vector using PCA and visualized each instance according to different emotions.

We applied this method to all participants and they showed similar boundaries between different emotions. Figure 4.7 shows an example of one subject's instances in five affective states, and different emotions are represented by different colors. Clear and discriminated boundaries can be observed between all the different emotions, indicating that the features we extracted can be used to distinguish between them. In Figure 4.7, each emotion type is one cluster with the same colors and symbols. There are five emotion clusters. Annoyance is purple crosses. Joy is purple left-pointing triangles pleasure is orange circles. Sadness is blue dots. Neutral is blue squares.

DTW-based Emotion classification

The extracted features were normalized to make the data comparable, and then fed into a DTW-based algorithm for emotion classification. The 20 time-sequenced features constructed one matrix 4.1 that is represented as one instance, where n in [1, N]. Then, equation 4.2 shows all the instances, where m is the instance number.

$$F_{instance=i}(n) = [feature_{1,instance=i}(n), feature_{2,instance=i}(n), ..., feature_{20,instance=i}(n)] \quad (4.1)$$

$$F_i(n) = [F_1(n), F_2(n), ..., F_i(n)] \quad i \in [1, m] \quad (4.2)$$

DTW allows non-rigid warping along the temporal axis and can also compensate for the

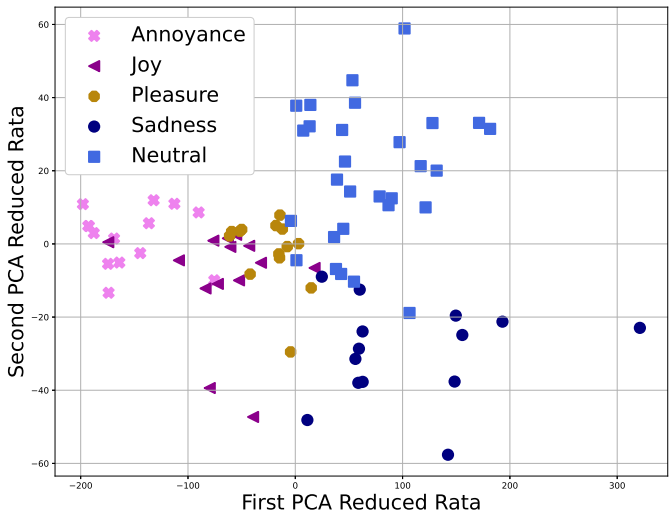


Figure 4.7: Emotion distributions of one subject's subject-dependent data by using PCA reduced features.

feature difference caused by different motion speeds [29]. Furthermore, DTW requires less training data and occupies less computational resources than other learning-based algorithms. Our DTW-based algorithm first selects templates for each emotion from the training data and then compares the testing instance to those templates. The instance is classified into the emotion templates with it has the shortest DTW distance. Specifically, each emotion's templates are the most representative instances of this emotion. They are selected by a pairwise comparison within each emotion during training. For every two instances, DTW distance is calculated between their feature sequences. The templates are selected for each instance based on whose DTW distances to all the other instances within the training dataset are minimal. The template number selected for each emotion can be tuned according to the number of users involved. We used 5 emotion templates in total in order to achieve both high performance and low computational cost.

The emotion template sets are constructed by these 5 instances as $F_i(n)|i = 1,...,5$. All types of emotion template are formed as $F_{i,emotions}(n)|k = 1,...,5; emotions = 1,...,5$. Then the testing data set will be compared with each type of emotion template 4.3.

$$
\begin{aligned}
Emotion_{predicted} = min[&\sum_{i=1}^{n} DTW(F_{test,emotion=1}(n), F_{train,emotion=1}(n)) \\
&, \sum_{i=1}^{n} DTW(F_{test,emotion=2}(n), F_{train,emotion=2}(n)) \\
&, \sum_{i=1}^{n} DTW(F_{test,emotion=3}(n), F_{train,emotion=3}(n)) \\
&, \sum_{i=1}^{n} DTW(F_{test,emotion=4}(n), F_{train,emotion=4}(n)) \\
&, \sum_{i=1}^{n} DTW(F_{test,emotion=5}(n), F_{train,emotion=5}(n))]
\end{aligned} \tag{4.3}
$$

4.5.4 Robotic Avatar Emotion Classification by Using CNN

The data on the movement of the robotic arm's joints are joint rotation angular trajectories across time. They contain spatiotemporal information, which means that the information captures both the spatial positioning of the robotic arm's joints in polar space and how these positions change over time. We employed the method [210] and encoded the spatiotemporal information contained within joint time sequences into 2D images. In specific, the joint sequences of the robotic arm are rotation angular values of six joints over time. The joint rotation time series are mapped into polar coordination, where the polar degree represents the joint rotation angles and the radius represents the time frames. We employed two steps to encode the spatial and time information into 2D images. The first step was to plot spatial information of seven joints as a curved line and the trajectories of the different 7 joints are depicted in distinct colours to illustrate their respective sequences. The second step was to encode the time information into the 2D image. The rotation angular value of each joint was plotted with a colour gradient from light to dark to present the temporal information.

One example encoded 2D image is illustrated in Figure 4.8. There are six different colored lines on the plot, each representing the rotational trajectory of a different joint over a time period. Each trajectory has a different fade color. Joint 0, 1, 2, 3, 4, 5 and 6 are represented by a grey, purple, blue, green, yellow and red line. The polar plots were scaled to 150 $\times$ 150 resolution image with no background grid before being input to the CNN.

Network Architecture

The structure of the CNN is built from three convolutional layers followed by four fully connected layers, as illustrated in Figure 4.9.

The convolutional layers are efficient in extracting high-level features in the input image and the dense (fully connected) layers flatten the features and make classification decisions. Each

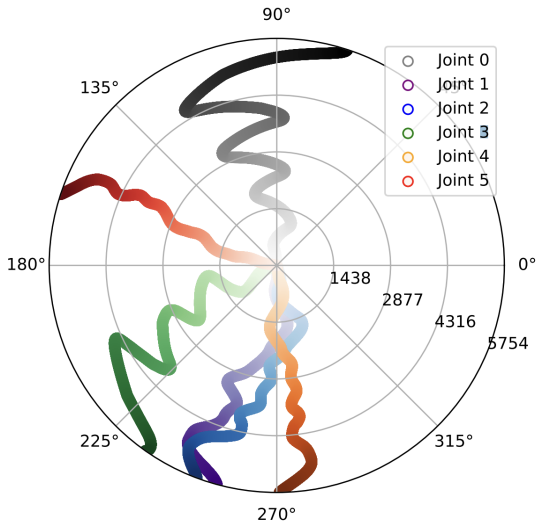


Figure 4.8: Six joint rotation angular trajectories mapping.

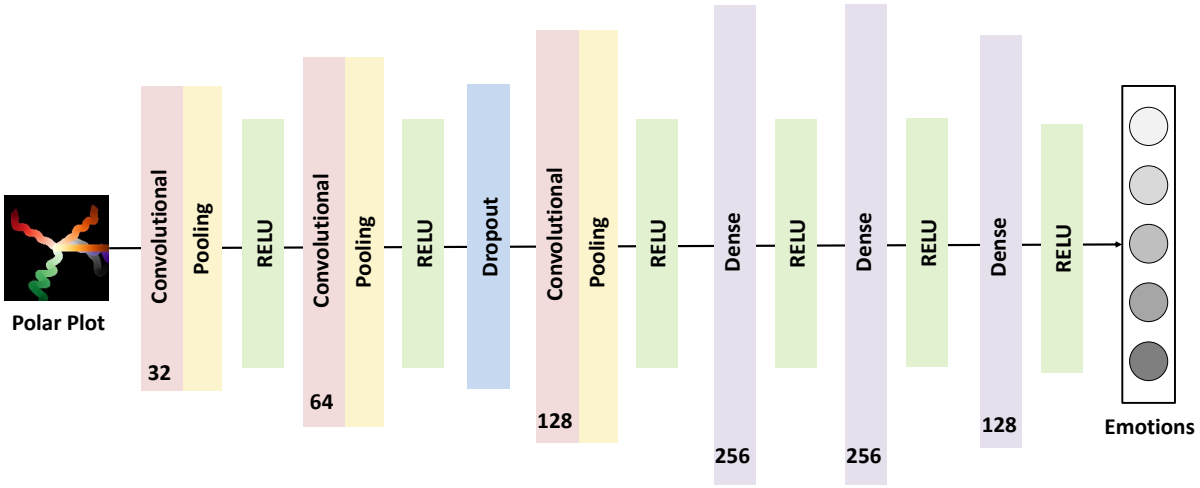


Figure 4.9: The CNN architecture starting from the left, the input is represented by an image of a polar plot. The network consists of several layers. Finally, the output is represented by a vertical bar labelled "Emotions".

convolutional layer is followed by max pooling and RELU functions. A dropout is added after the second convolutional layer to prevent overfitting. The size of the final dense layer has the same output size as the number of emotions. The pseudo-code of the algorithm is given below 3. Firstly, we processed the input and output data. Then, we built the network model, shown as illustrated in Figure 4.9. Then, we trained the model. We employed the cross entropy as the loss function that is . We used the RMSProp as the optimization function that is .

In order to understand the number of parameters in convolution neural networks (CNNs), we calculated the parameters of the different layers shown in 4.2.

Algorithm 3: Pseudo-code of CNNs Model

Input: X ← Robotic joint Images in 150 × 150× 3 dimension
Output: Y← Categories (Annoyance, Pleasure, Sadness, Joy and Neutral states)
X: X← Normalization (X)
Y: Y← Encoding(Y) (Annoyance=0, Pleasure=1, Sadness=2, Joy=3 and Neutral=4)
//..... *Build CNNs Model* .....
Model ← Sequential()
Model← CNN layers
//.....
//..... *Performing training and testing of CNNs* .....
kf=KFold(n_splits=1)
For *train index, test index in kf.split(X)* **do**
x train← X[train index], *y train*← Y[train index]
x test← X[test index], *y test*← Y[test index]
model.compile(loss="CrossEntropyLoss"; optimizer="RMSprop", metrics=["accuracy"])
history←model.fit(*x train, y train*, epochs = 40)
Calculate performance metrics(confusion matrix and accuracy) for each epoch
End for
Output CNNs classifier and classification results

Table 4.2: Parameters in Convolution Neural Networks (CNNs).

	Activation Shape	Activation Size	Parameters
Input Layer:	(150,150,3)	67500	0
CONV1 (f=3, s=1)	(148,148,32)	700928	896
POOL1	(74,74,32)	175232	0
CONV2 (f=3, s=1)	(72,72,64)	331776	18496
POOL2	(36,36,64)	82944	0
CONV3 (f=3, s=1)	(34,34,128)	147968	39296
POOL3	(17,17,128)	36992	0
FC4	(256,1)	256	9470208
FC5	(256,1)	256	65792
FC6	(128,1)	128	32896
FC7	(5,1)	4	645

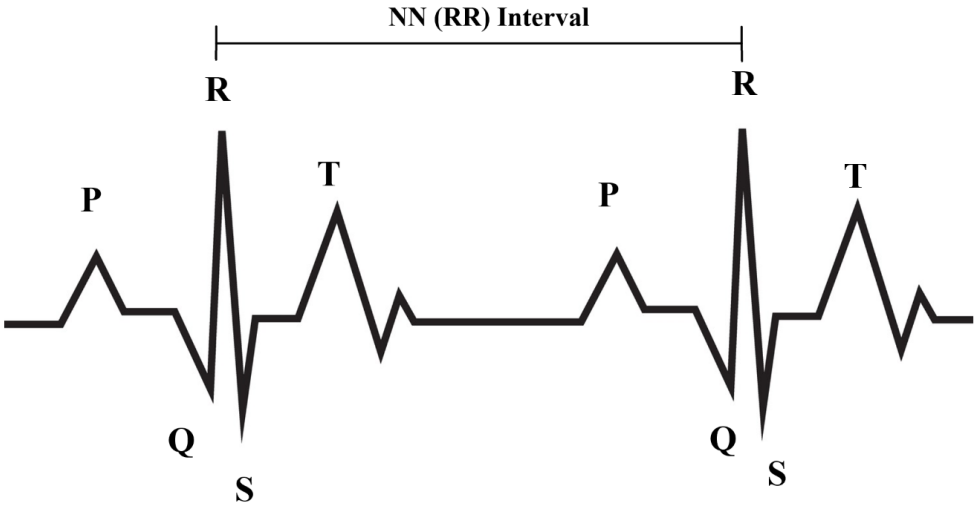


Figure 4.10: A standard ECG signal.

4.5.5 ECG Signals Emotion Classification

We selected ECG as a comparable emotion recognition method due to its potential suitability for the task, when compared to other options such as facial or voice. Changes in facial expressions or voice signals may be hard to capture in a telerobotic scenario, as this would require head-facing cameras to be installed in telerobotic workstations, while speech is not inherent to many tasks. Besides, ECG is a well-established and state-of-the-art emotion classification sensing method [60] and shows higher performance compared with other physiological signals on the emotion recognition reaching 90.0% [68]. An ECG detects the electrical activity of the robot operator's heart in real time [60]. Each heartbeat is a specific waveform, QRS complex waveform, caused by the ventricles contraction 4.10 [60, 104, 228]. For each beat, there are 6 main points (P, Q, R, S, T, U) and the NN interval refers to the time between two R-peak, shown as Figure 4.10.

We then extracted emotion features from time-domain and frequency-domain features, including pNN50, Welch PSD: LF/HF, Lomb-Scargle PSD: LF/HF, Autoregressive LF/HF, Poincar SD1, Poincar SD1/SD2, and Detrended Fluctuation Analysis(DFA), following well-established methodology from across prior work [60, 104, 105, 228]. These features are described in 4.3. For DFA, X_t is divided into windows of different lengths, with the window length denoted as n, and then the squared error within each time window is minimized to obtain a fitted line of local least squares, where Y_t is the fitted line. Lomb-Scargle PSD computes a Power Spectral Density (PSD) estimation from the NNI series using the Lomb-Scargle Periodogram. Lomb-Scargle PSD LF/HF is the ratio of high frequency band and low frequency band. Autoregressive computes a PSD estimation from the NNI series using the Autoregressive method and autoregressive PSD LF/HF is the ratio of high frequency band and low frequency band. Then these features are fed to the Support Vector Machine (SVM) [228] for emotion classification. We selected the linear kernel which is the most efficient one of the SVM. SVM is a supervised learning algorithm

Table 4.3: ECG features extraction for KNN.

Features	Formula
pNN50	PNN50=$\frac{NN50}{TotalNN\times 100\%}$;
Welch PSD: LF/HF	$\mathrm{P}_{x_m,M(w_k)} = \frac{1}{M}\lvert\mathrm{FFT}_{N,k}(x_m)\rvert^2 = \frac{1}{M}\lvert\sum_{N-1}^{n=0} x_m(n)e^{-j2\pi nk/N}\rvert^2$
Lomb-Scargle PSD: LF/HF	Ratio of LF and HF of PSD using Lomb-Scargle method
Autoregressive: LF/HF	Ratio of LF and HF of PSD using Autoregressive method
Poincar SD1	$x = RR1, RR2, ..., RR_n$ $\mathrm{meanRR}_n = abs(RR_n - RR_{n+1})$ $SD_1 = \sqrt{0.5(std(RR))^2}$
Poincar SD1/SD2	$SD_2 = \sqrt{2(std(RR))^2 - 0.5(std(RR))^2}$
Detrended Fluctuation Analysis(DFA)	$X_t = \sum_t^{t=1}(x_i - \langle x\rangle)$ $F(n) = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(X_t - Y_t)^2}$

designed for classification, regression, and outlier detection tasks. SVM particularly is efficient in high-dimensional datasets.

4.6 Experiments

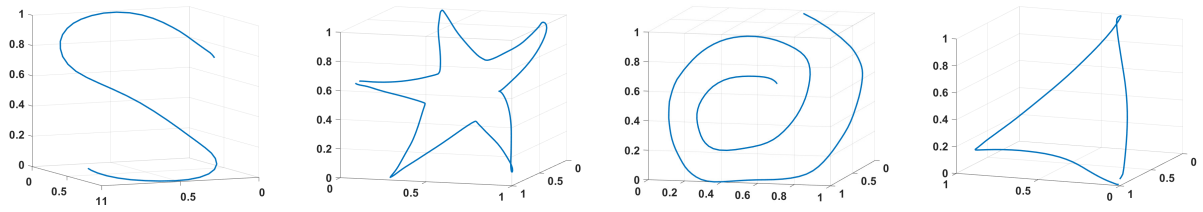
In this section, we describe the experiments conducted on the motion-controlled robotic avatar platform to evaluate our emotion classification approach, including emotion stimulation, motion selection, and experimental setup.

4.6.1 Emotion Stimulation

Emotion stimulation is an established method for obtaining high-quality emotion data [228]. People have different emotional reactions to sound and music based on their personal experiences, and cultural background and musical training can also influence these reactions [104, 228]. Given this, we followed an established personalisation approach: participants were allowed audio files of their choice, such as songs, noise, or even stand-up comedy recordings to stimulate each corresponding emotion individually [104, 228, 228]. While choosing not to strictly control the emotional stimuli could add variance, we chose the personalised approach in the absence of any truly consistent way to elicit emotion in people, a limitation which impacts the entire field.

4.6.2 Non-Stylized Motions

Our work focuses on the functional, non-stylized motions that are likely to be performed during telerobot control, rather than motions intended to express emotion as the primary goal [44]. Specifically, we designed 14 non-stylized motions as tasks and divided them into two categories. The first category was mid-air gestures, which are performed in many motion-controlled robotic



(a) "S" (b) "Star" (c) "Stir" (d) "Triangle"

Figure 4.11: Designed tasks with curve lines, straight lines, and sharp curve characteristics. There are four 3D trajectory figures drew by robotic arm including one s letter, star, stir and triangle.

avatar scenarios to move the perform actions in the 3D space. The second category is the line-tracing tasks, commonly performed in motion-critical scenarios. Operators are required to move the robotic avatar to follow pre-designed trajectories, which restrict their motions.

Mid-air Gestures

We designed nine mid-air gestures: (1) cursive "Lw", (2) "Star", (3) "Stir", (4) "S", (5) "Triangle", (6) "Drinking", (7) "Knocking", (8) "Throwing", and (9) "Waving". The mid-air gestures (6) to (9) are examples of typical social tasks that appear in social HRI and freestyle tasks in remote education scenarios [204, 260]. The participants controlled the robotic avatar to perform these mid-air gestures in a non-prescriptive manner without hard constraints.

Line-tracing Tasks

We also designed five line-tracing tasks which contain typical features of motions that would appear in motion control scenarios including (1) "Lw", (2) "Star", (3) "Stir", (4) "S", and (5) "Triangle" (see Figure 4.11). For example, "Star" and "Triangle" contain sharp turning points. "S" contain smooth turns, while "Stir" contains consecutive turns. Cursive "Lw" synthesizes the all features of the other drawing tasks. Compared with the mid-air gestures, the line-tracing tasks required the participants to follow the printed trajectory reference, such as robotic-assisted spine surgery [207], total knee arthroplasty [261] and dental implantology [262]. To be more specific, we add constraints to these tasks, which simulate mission-critical control scenarios.

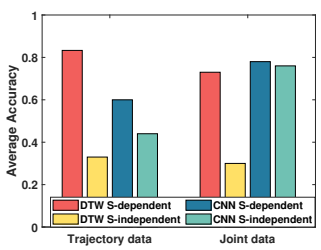
In this work, we followed an approach of Huang et al.: utilising a classification of foundational tasks that can be widely applicable to different telerobotic scenarios [225]. We included a selection of task types representative of the core movements of real-world tasks, including surgery, education and social scenarios, as featured in prior work [204, 207] to provide a good foundation. However, our study has a necessarily narrower scope than real-world use, which should addressed in future by co-designing tasks with real-world telerobotic users, to ensure specific and highly ecologically valid tasks for specific domains.

4.6.3 Experimental Setup and Data Collection

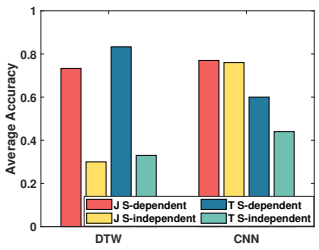
We recruited ten volunteers (3 female, 7 male) to conduct the experiments. They are all university students with ages ranging from 20 to 25 (mean = 24.3, σ = 1.42). All ten volunteers first read an information sheet, signed a consent form and brought their individual pre-pickup audio clips (See Section 4.6.1) to trigger emotional responses (joy, pleasure, annoyance and sadness). Before the experiments, each volunteer underwent 30 minutes of training to become familiar with the control process. Then, participants took a 10-minute break to relax and calm down, which was done to help them reset to a more neutral emotional state. The experiments were conducted using the motion-controlled Universal Robot UR3e platform, as introduced in Section 4.3.1, which is deployed in a quiet $50m^2$ laboratory room. Six OptiTrack cameras are placed in a $5m$-by-$5m$ square area. Participants wore a marker glove on their right hand and the ECG devices were attached to their other hand and the ipsilateral ankle. During the experiments, participants sat in the centre of the OptiTrack cameras to control the UR3e while listening to the audio they selected. Operators were given three minutes of emotional stimulation at the beginning of each task and the collection time was approximately 2.5 minutes per task. A 10-minute emotional recovery break was given between each task. The order of emotion stimulation was random to offset the influence between different emotions.At the end of each task, an interview was conducted to assess if the stimulated emotions were consistent with the target emotions. If yes, then the collected data was labelled with the subject's reported emotion. As "Lw" synthesized all the drawing motions task, we required all participants to perform both the line-tracing "Lw" and in-air "Lw" under all five emotions (joy, pleasure, sadness, annoyance, and neutral). Each task is performed 15 times under each emotion. Given this, there were 1500 total instances of the "Lw" task (10 subjects $\times$ 5 emotions $\times$ 15 times $\times$ 2 non-stylised tasks types). Then, five of the ten participants are asked to perform the remaining twelve tasks, which include eight mid-air gestures ("Star", "Stir", "S", "Triangle", "Drinking", "Knocking", "Throwing" and "Waving") and four line-tracing tasks ("Star", "Stir", "S" and "Triangle"). These five participants performed each task 15 times under each emotion, resulting in 4500 (5 subjects $\times$ 5 emotions $\times$ 15 times $\times$ 12 non-stylised tasks types) task instances. We did not ask the remaining 5 participants to complete these extra tasks due to time constraints per experimental session, as these participants took significantly longer to complete the "Lw" task. In total, 6000 task instances were collected from the experiments to serve as the dataset.

ECG data collection

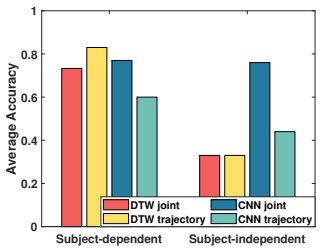
As mentioned in Section 4.2.3, ECG is a well-established emotion recognition method, so we collected ECG data to compare its suitability to our method. There are reasons why ECG may not be ideal for telerobotic scenarios: ECG signal collection for emotion classification is normally taken for around 8 minutes [228] [104], and it can be confounded by motion artefacts. In telerobotic motion control scenarios, the subject is not liable to be stationary and tasks can



(a) Trajectory and joint data comparison



(b) DTW and CNN algorithm comparison



(c) Subject dependent and independent model comparison

Figure 4.12: The emotion classification results for different classifiers trained by different algorithms and different data ("S" stands for "Subject", "J" stands for "Robot Joint Data", and "T" stands for "Robot Trajectory Data").

be quite short, such as pick and place or stirring tasks [215, 263, 264]. Even in longer tasks, such as teledriving, the inherent motion of the task has led to ECG classification being limited in past work [265–268]. We sought to understand whether our proposed approach could prove a more suitable alternative, and so compared the two in this work. To assess this, we outfitted participants with an integrative ECG [249] Attys biomedical sensor device on the hand they did not use for operation and instructed them to keep this hand as steady as possible. The ECG signals collected during task execution were then used to implement traditional ECG-based emotion classification, to compare its suitability with our proposed robotic avatar emotion classification approach.

4.7 Results and Analysis

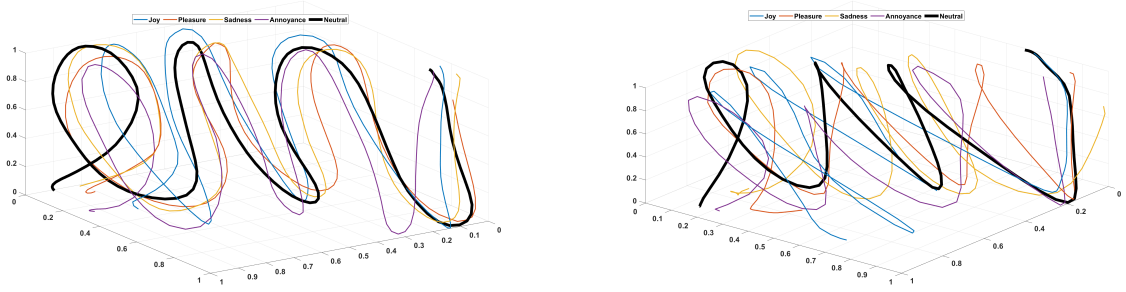
In total, 6000 instances were collected from the UR3e platform to evaluate our emotion classification approach on the robotic avatar, a relatively large amount of data when compared to prior work [44, 228]. Following the task classification [225], we trained two types of emotion classifiers for each task category: a subject-dependent classifier and a subject-independent classifier. The subject-dependent classifier was trained and tested on the dataset of each individual subject, while the subject-independent classifier was trained in a leave-one-subject-out procedure.

Emotion expressions are highly individual [269] and vary between people. Subject-dependent classifiers can study these individuals' emotional expressions, allowing for more accurate and sensitive emotion detection by providing high personalization. This personalised training and optimisation could be deployed to individuals when using shared workplace equipment when they identify themselves by logging in. Subject-independent classifiers are, however, also important and a priority for real-world use, as it is likely that classifiers would also be trained on the common emotion information of other operators. Thus, we implemented and explored both subject-dependent and independent classifiers to fully explore this field.

We used a DTW-based algorithm (introduced in Section 4.5.3) and a CNN-based algo-

rithm (introduced in Section 4.5.4), respectively, to train the subject-dependent and subject-independent classifiers. The emotion classification results for different classifiers trained by different algorithms and different data are shown in Figure 4.12. There are three bar graphs and each bar graph has the result of accuracy of different algorithm and datatype combinations. The first graph compares the average accuracy of two algorithms, DTW and CNN, with the data being either subject-dependent or subject-independent. The bars are color-coded to distinguish between algorithms and dependency, with four sets of bars representing the accuracy of each condition applied to either trajectory data or joint data. The second graph also displays average accuracy but focuses on comparing the performance of the DTW and CNN algorithms under the conditions of joint and trajectory data, with each being subject-dependent or subject-independent. The third graph, shows the average accuracy of DTW and CNN algorithms when applied to joint and trajectory data, with distinction between subject-dependent and subject-independent. The bars are grouped by subject condition and then further by the type of data with each algorithm's performance distinctly color-coded. In particular, Figure 4.12a shows that, when using the robot end-effector's trajectory data as the input, the DTW-based algorithm achieved the highest performance, while when using the robot joint data as the input, the CNN-based algorithm had the best performance. Figure 4.12b shows that the average performance of the CNN-based algorithm is better than the DTW-based algorithm, as the DTW-based algorithm did not perform well in training the subject-independent classifier. Figure 4.12c shows that the accuracy of subject-independent classification is lower than that of subject-dependent classification, which is as expected. Based on the above results, we decided to use the DTW-based algorithm and the robot end-effector's trajectory data to train the subject-dependent classifier and use the CNN-based algorithm and the robot joint data to train the subject-independent classifier. Individual differences in emotion expression may explain the lower performance of the DTW algorithm for subject-independent data. For example, one person may move faster or further than another, even when both are annoyed. The DTW algorithm measures the distance difference between two different emotional instances, so different expressions of emotions between people make the emotion instances less comparable and finding common emotion information across people more difficult. The emotion classification performance of these two types of classifiers is presented below.

The trajectories of the performed task under different emotions are shown in Figure 4.13. There are two 3D letter lw trajectories. Each 3D box contains 5 lw trajectories and different colors are used to represent different emotions. The left one is the Line-tracing lw and the right one is the mid-air lw.



(a) Line-tracing "Lw". (b) Mid-air "Lw".

Figure 4.13: The trajectory of line-tracing "Lw" task and mid-air "Lw" task under different emotions.

4.7.1 Subject-Dependent Results

Classification Performance Variance By Subject

Figure 4.14 is a bar graph showing the accuracy of each subject. The provided image is a vertical bar graph representing the average accuracy for two different tasks across a sample of ten subjects. Each subject is represented by an index number on the x-axis, which runs from 1 to 10. The y-axis measures the average accuracy, ranging from 0 to 1. There are two sets of bars for each subject, color-coded to distinguish between the two tasks: mid-air gestures and line-tracing tasks. It shows the average emotion classification result of all mid-air gestures and line-tracing tasks, for each of the ten subjects respectively. Overall, the average accuracy achieved by the subject-dependent classification among all subjects was 83.3%. We can observe that the performance on mid-air gestures and line-tracing tasks are similar among different subjects, which indicates the proposed approach works for different types of motions. The results show our approach can classify each operator's emotions with relatively high accuracy.

Classification Performance Variance By Task

The subject-dependent classifier's average performance among all the subjects for each of the mid-air gestures and line-tracing tasks is presented in Table 4.4. The average accuracy achieved by the mid-air gestures was 86.5%, while the average accuracy of the line-tracing tasks was 77.9%. The performance of the ten drawing tasks ranged from 73.7% to 92.0%, with the "S" task showing the best overall performance in both the mid-air gestures and line-tracing tasks.

Table 4.4: Subject-dependent emotion classification for each of the mid-air gestures and line-tracing tasks.

Tasks	Lw	Star	Stir	S	Triangle	Drink	Knock	Throw	Wave
Mid-air gestures	0.851	0.853	0.817	0.913	0.920	0.807	0.860	0.858	0.909
Line-tracing tasks	0.777	0.737	0.757	0.833	0.791	NA	NA	NA	NA

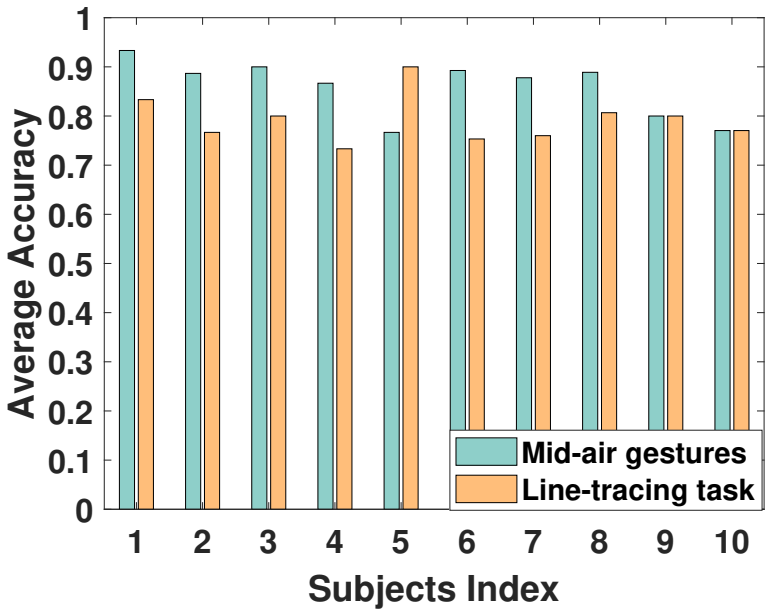


Figure 4.14: Subject-dependent emotion classification of mid-air gestures and line-tracing tasks for ten subjects.

The performance of the four social tasks ranges from 80.7% to 90.9%, with the "Wave" task performing the best. The results show that our subject-dependent algorithm can be used to infer emotions from a wide range of tasks performed by different users.

Classification Performance Variance By Emotions

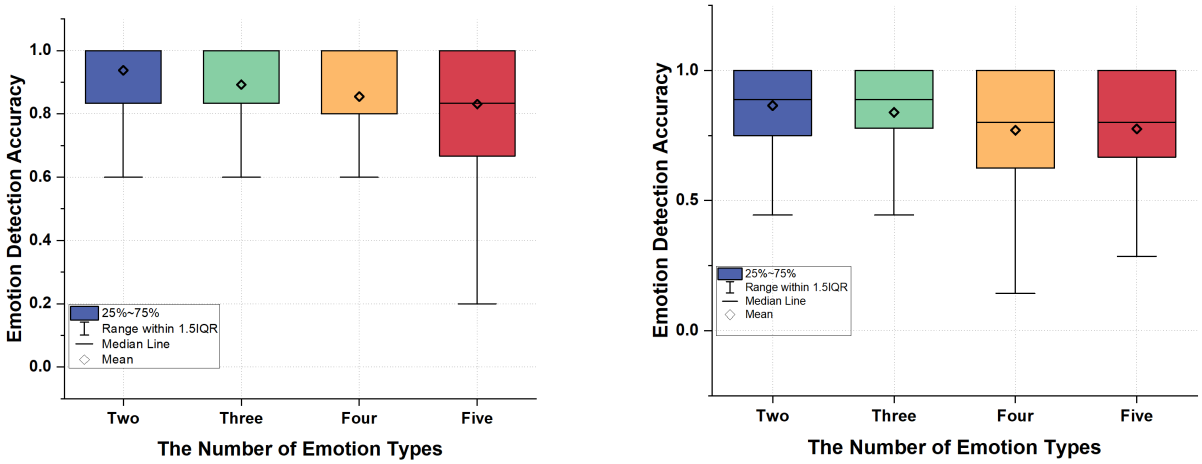
The average detection accuracy for each type of emotion among different tasks performed by each subject is presented in Table 4.5. Our approach detected all five types of emotions with an average accuracy of 83.3%. In particular, the average detection rate for "Joy", "Sadness", "Annoyance", "Pleasure" and "Neutral" among the ten subjects is 83.12%, 86.67%, 90.75%, 68.11% and 87.31%, respectively. The results show that this approach generally works for detecting different types of emotions. Pleasure was something of an outlier with worse performance, which may indicate that being in this high valence low arousal affective state resulted in less distinct and expressive movement features than the other emotional states, particularly Joy (high valence high arousal).

Classification Performance Variance By Number of Emotions

Figure 4.15a is subject-dependent emotion accuracy result among 2, 3, 4, and 5 emotion types using DTW method. It shows the average emotion classification results among all tasks and subjects when different numbers of emotion types are involved. Each box plot is color-coded differently and contains a diamond shape that represents the mean accuracy. The boxes themselves represent the interquartile range (IQR), extending from the 25th to the 75th percentile,

Table 4.5: Subject-dependent classifier's average emotion detection accuracy for different subjects.

Subject No.	Annoyance	Pleasure	Sadness	Joy	Neutral
1	0.9167	0.7500	0.9167	0.9167	0.9167
2	0.8488	0.8750	0.8310	0.7952	0.8690
3	1.0000	0.5500	0.9000	0.9000	0.9000
4	1.0000	0.3333	0.9167	0.9167	0.8333
5	0.9167	0.8333	0.7500	0.6667	1.0000
6	0.8810	0.7381	0.8810	0.8571	0.8571
7	0.9286	0.7619	0.8452	0.7024	0.9405
8	0.8214	0.7452	0.9262	0.8762	0.9286
9	0.9167	0.5000	0.9167	0.9167	0.7500
10	0.8452	0.7238	0.7833	0.7643	0.7357
Average	0.9075	0.6811	0.8667	0.8312	0.8731
Standard deviation	0.0608	0.1674	0.0621	0.0936	0.0832



(a) Subject-dependent classifier.

(b) Subject-independent classifier.

Figure 4.15: Emotion classification results regarding different numbers of emotion types for different classifiers.

with 'whiskers' that indicate the range within 1.5 times the IQR above and below the box. A horizontal line within each box shows the median accuracy. The second box plot, Both plots have an x-axis that categorizes the boxes by the number of emotion types, and a y-axis that measures emotion detection accuracy from 0 to 1. When classifying two types of emotions (*e.g.*annoyance and neutral), our approach achieved an average accuracy of 94.05%. When classifying three types of emotions (*e.g.*joy, annoyance, and neutral), the average performance degraded to 89.8%. And when classifying four types of emotions across all types, the performance further degraded to 86.3%. When there are five emotions (joy, pleasure, sad, annoyance, and neutral) to be classified, our approach can still achieve an accuracy of 83.3%. The results demonstrate that our approach is able to classify common emotions with relatively high accuracy.

4.7.2 Subject-Independent Results

Trajectories of a task performed by different operators under the same emotion showed different motion features, thus it is hard to use the robot end-effector's trajectory to implement subject-independent emotion classification, as Figure 4.12 shown. Instead, we use the robot joint data to train the subject-independent classifier introduced above. Intrinsically, subject-independent emotion recognition is more of a challenge as the way subjects express emotions varies, which can be seen manifesting in the performance variance of the DTW (see Section 4.7). As expected, subject-independent performance was lower than that of subject-dependent. We split the datasets into 60% training data and 40% testing data. Across tasks and five affective states, the average accuracy of emotion recognition on testing data was achieved at 74.2%.

To study the capability of our system to recognize the emotions of users not included in the training model, we utilised a leave-one-subject-out cross-validation (LOSOCV) procedure(each subject was left out of the training dataset for their testing), as discussed in Section 3.2.2. The pseudo-code of LOSOCV is shown in 4.

Algorithm 4: Pseudo-code of LOSOCV Algorithm

1. Divide data into N folds ▷ N is the number of instances
2. **Loop** For 1 in N:
    1. Set 1 as the testing set
    2. Set N-1 as the training set
    3. Train model on N-1 training set
    4. Test model on 1 testing set
3. Calculate the average performance over N

Across tasks and affective states, the LOSOCV analysis achieved an average emotion recognition accuracy of 76.5%. The following sections discuss accuracy variance between tasks and affective states.

Table 4.6: Subject-independent emotion classification for each of the mid-air gestures and line-tracing tasks.

Tasks	Lw	Star	Stir	S	Triangle	Drink	Knock	Throw	Wave
Mid-air gestures	0.775	0.734	0.757	0.758	0.828	0.711	0.789	0.719	0.844
Line-tracing tasks	0.682	0.828	0.664	0.766	0.852	NA	NA	NA	NA

Table 4.7: Average emotion detection accuracy achieved by the subject-independent method.

Emotion	Annoyance	Joy	Sad	Pleasure	Neutral
Detection Accuracy	0.676	0.779	0.870	0.700	0.779

Classification Performance Variance By Task

The subject-independent emotion classification result for each of the mid-air gestures and line-tracing tasks is shown in Table 4.6. The "LW" task was trained and tested on 10 subjects while the rest were trained and tested on 5 subjects. Our approach achieved an average accuracy of 76.8% and 75.8% for the mid-air gestures and line-tracing tasks, respectively. The emotion classification performance ranges from 66.4% to 85.2%, which is comparable to the results of the existing work [44] that used human motion signals for emotion classification.

Emotion Detection Accuracy

As shown in Table 4.7, our subject-independent method achieved over 76.5% emotion detection accuracy across all five types of emotions: "Joy", "Sadness", "Annoyance", "Pleasure", and "Neutral". Interestingly our approach achieved higher detection accuracy for emotions with lower arousal. A possible reason for this could be that low arousal emotions may contain more subject-independent features than high arousal states.

Classification Performance Variance By Number of Emotions

Figure 4.15b is subject-independent emotion accuracy result among 2, 3, 4, and 5 emotion types using CNN method. The figure shows similar information as the Figure 4.15a. It shows the performance of our subject-independent method when there are different numbers of emotion types. Specifically, the average emotion classification accuracy is 88.3% for two emotions (*e.g.*, annoyance and neutral), 84.0% for three emotions (*e.g.*joy, annoyance, and neutral), 80.9% for four emotions (*e.g.,* joy, sad, annoyance, and neutral), and 76.5% for all five emotions (joy, pleasure, sad, annoyance, and neutral).

4.7.3 Our approach versus ECG-based emotion recognition

As we discussed in Section 4.2.3, ECG-based emotion classification uses minor changes in physiological signals (heart rate) to detect emotional changes and requires physical contact sensors to

observe the ECG signal. These two factors constrain the application scenarios of the ECG-based emotion classification. In order to compare the suitability of our emotion recognition method, we collected the ECG data of the subjects during the study. The time length of each ECG instance was around 210s. We used IIR to filter the ECG data and extracted the emotion-related features as mentioned in 4.5.5. As expected, the average accuracy across ten subjects was 56.6% lower than the existing results 87% [228], likely due to both motion artifacts and measurement duration.

4.8 Discussion

Our work shows that a robotic avatar's motion behaviors can be used to infer the operator's emotions. We know now that emotions can have a vital role in the interaction between humans and robots, as they have a direct impact on the control of the remote robot. Thus, it is beneficial to understand and observe operator emotions to avert any erroneous operations that may potentially cause harm to the safety-critical scenarios.

4.8.1 Current Performance of the Approach, Limitations and Next Steps

Current performance of the Approach

Emotional Information Involved in Interaction

Our robotic avatar can inherit human hand trajectories but can not reproduce human trajectories perfectly. On the one hand, the skeleton and the degree of freedom (DOF) of the robot arm and the human arm are different. On the other hand, the limitation of the control algorithm and the communication delay cause a deviation between the robot's and human's trajectory. Although the robotic arm can only reproduce lossy trajectories, our emotion classification outperforms the work [44] of using individual human status data. The data set for their work [44] was 235 and the average classification performance among four emotions for five subjects was 70.05%. The data set of our work is 6000 and the average classification performance among five emotions for five subjects is 83.3%. This indicates that measuring emotion using our methodology via inference of robotics arm trajectories may be as, or more, sensitive than prior approaches. Both our work and Loghmani et al. [44] used non-stylized motions, but our participants were performed in an interactive control scenario, i.e., operators observe a robot's movements to adjust their own behaviours in real-time. It could be that emotional expressions are more pronounced in such interactive control scenarios. The influence of interaction scenarios on emotion expression should be further explored in the future, which could profoundly impact subsequent interaction design.

Benefits of Telerobot Emotion Classification Compared to Traditional Methods

We attached the ECG device to the operators' stable hands and ankles to capture their heart

rate signals. We found this to be a limitation of the ECG experiment setup, which requires longer measurement time and for operators to remain stable for optimal performance, as in motion-controlled scenarios, users need to move and are unlikely to remain still enough. Similarly, many existing methods of using physiological signals to classify emotions require humans to stay stable, which limits the application of ECG emotion classification. While applicable in controlled laboratory experiments, these limitations would preclude the real-world use of these techniques in telerobot scenarios. Our work verifies, for the first time, the limitation of ECG in human remote-control robot scenarios and shows that this method lacks ecological validity for this use case. For example, when we evaluate whether a driver is exhausted or not, we can not require him/her to stay stable while driving. Our approach overcomes this limitation, however, by utilizing their motions to infer emotions. From this point of view, behaviour-based emotion classification is more practical and promotes many applications in this field.

Limitations and Future Work

In this study, we used audio files, a well-established emotion stimulation method [270], to elicit emotions and interviewed participants to check they felt the correct emotion was evoked. This approach has limitations, however, as it has been regarded as non-immersive when immersion is an important aspect of eliciting emotions in real experiences [270]. Future work could seek to adopt a more immersive emotion elicitation approach, such as leveraging Virtual Reality [270].

Another core limitation of this work is that emotion is inherently ambiguous and complex, so there may exist disagreements between participant annotators' labels and their real emotions [269]. In addition, while we took steps to help participants regain a neutral emotional state between tasks, this cannot be fully controlled. While we used established methodology in this work, this is a general problem within the field of affective computing [105].

Our work features a participant sample size of 10, with some tasks only performed by 5, which limits the immediate generalisability of our current model to the wider population and real-world applications. We did, however, collect 6000 instances in total, a larger set than similar prior affective computing studies [44, 228], and we achieved comparable results. This demonstrates the feasibility of our emotion inference method, but in future, a larger participant pool conducting a wider set of tasks would help in directly applying this method to remote-operation scenarios. Emotional expression may also vary in intensity in different tasks. For example, linear motions may show less emotive features than complex motions. This should be accounted for when aiming to achieve real-world generalisability. Similarly, future work could adopt differentiation between tasks used in training and testing to further explore applicability to unseen real-world tasks. Participants also received 30 minutes of training, which is necessarily limited compared to a full training regime for real-world telerobotic operation.

In this work, we used an ECG device, but found it unsuitable for telerobotic scenarios, as ECG data collection suffered from motion artifacts in the data. This intrinsic limitation made

it difficult to capture effective emotional information, and thus it was difficult to compare its efficacy with our method. Future work should seek to utilise additional sensors for multimodal data, such as EEG and respiration rate, which could achieve accuracy comparable with that previously shown by ECG [68] while minimising the impact of motion artefacts [89]. This would allow for a more robust comparison with this novel robot motion-based inference approach.

4.8.2 Implications for Current and Future Telerobotic Applications

Current Remote Robot Operation:

Our work found that operator emotion can be successfully inferred from telerobotic movement and that certain emotions can result in more vigorous and pronounced movement. With this in mind, it is prudent to consider how this might impact current telerobot applications differently. For example, telerobotic keyhole surgery is an extremely precise and safety-critical environment where small movements could have dire health consequences. Given this, intervening swiftly to remove control during moments of heightened operator emotion could be highly beneficial. This would render the end-effector suddenly stationary, which is unlikely to be consistently dangerous, as keyhole surgery is made up of prolonged pauses and slow movements, but could be problematic if the effector is currently interacting with tissue. Furthermore, keyhole surgeons are highly specialised, so handing over control to a replacement operator may be difficult.

By contrast, teledriving presents a more difficult scenario for intervention after knowing the driver's emotions. While prior work has observed driver emotion directly from onboard control telemetry [96], it is still unclear how this information should be applied to reduce danger or risk. While driving is a less precise task than telesurgery, it is still a safety-critical task where erratic operator behaviour may warrant the removal of control. Unlike telesurgery, however, removing control of teledriving leaves a vehicle that is still in motion, potentially turning and will need to be brought to a controlled stop. Given this, an intervention may need to either hand over to an autonomous driving system or perhaps reduce noise in the operator's control, rather than remove it (see 4.8.2). Another context to consider is industrial applications, such as the telerobotic handling of nuclear waste containers. While also safety-critical, this operation is less precise than surgery, making the emotional level required to intervene more extreme, and control may be safely paused and handed over to another remote technician to complete the task. Beyond the differing practical concerns of observing operator emotions and intervening in different contexts, we must also consider the potential impact on these humans in the loop and how the system can be designed to be cooperative with users, rather than combative.

More enhanced feature extraction and advanced machine learning techniques are required in the future to generalise tasks across varied operation in the real world. For example, transfer learning-based methods could be developed [271–273] that can transfer previously learned knowledge from available large-scale data and establish a new model. We could utilize such

approaches to achieve cross-subject emotion recognition. Outside of using motion tracking to control remote robots using the Optitrack sensor, alternative control schemes, such as haptics globes and physical controllers have been applied to control remote robots. It would be valuable to explore if our method of emotion inference is still effective across these input devices through formal testing and investigation.

Understanding the Human Impact:

While emotional inference could be used to intervene during safety-critical telerobotic scenarios, possible pitfalls must be considered. First is the issue of privacy. Safety-critical telerobotic operation supports various applications including those for which humans cannot be physically present, such as nuclear waste handling, as well as healthcare, transport and industry. Working remotely can afford employees increased privacy when compared to those to work on-site, which they may value [274,275]. Operators may feel that having their emotional state inferred through robotic arm movement infringes on their right to work while managing their private internal emotional state. While co-located workers would naturally display emotional cues through their body language or voice, systematically monitoring and using their emotions to assess performance or intervene for safety reasons would require fitting with traditional electric-signal-based monitoring devices. As discussed in Section 4.2.3, such devices can be confounded by the movement inherent in telerobotic operation. Our system could, therefore, offer a functional replacement for this context. While it would be clear to an employee that they have been fitted with a wearable monitor, it may be less clear that they are being monitored based on robotic avatar movement. Thus, this system should be clearly signposted and the informed consent of operators obtained.

Another issue operators could experience is fear of loss of agency, as they know their control could be removed due to automatic inference of their involuntary emotional state, which could in turn lead to an adversarial relationship between user and system. For example, operators may seek to practice emotion regulation using real-time Response Modulation [276] in order to avoid losing control, which in turn may deplete attentional resources and risk worsening performance. Losing control, when they otherwise would not have, could also damage an operator's confidence and mental well-being. If the loss of control is observable by peers, it may also lead to perceptions of incompetence or feelings of shame. Repeated interventions or interruptions by such a system could also be seen as frustrating or annoying. Given this, such a system should be implemented in an ethical and empathetic manner, with the removal of control treated as a safety-driven last resort, in order to mitigate users harbouring resentment for the system. As an aside, there are less disruptive ways emotional information during interaction could be used, such as evaluating the operator satisfaction as feedback to improve the robot's control algorithms. In the next section, we propose an alternative moderate approach to removing operator control, emotive-motion dampening, which could mitigate these issues while still improving

safety outcomes.

Future Applications:

AI-Assisted Emotive-Motion Dampening

As discussed, emotional influence on telerobotic avatar movement could have negative safety outcomes, but simultaneously the sudden removal of operator control based on their emotional state could have negative practical and psychological ramifications. Given this, we propose an intermediate solution, the real-time dampening of emotive-motion features. When enabled, real-time AI would be leveraged to filter out the drastic and jerky features of user input motion that are caused by a high-arousal emotion state, normalising to a smoother trajectory. This approach is analogous to the aim-assist feature used in some first-person video games [277], or with *shared control* paradigm explored in prior work [278–280], whereby control is shared between the human and the robotics autonomy. Extending prior work, we propose to apply this technique responsively based on the operator's inferred emotional state.

In some scenarios, this could be enabled by default, although in high-precision scenarios, such as telesurgery, it could reduce the operator's level of fine-grain control. In these scenarios, the dampening system could instead be enabled only when an emotional state which could compromise the safety-critical task is detected. Such a system could also have privacy benefits, as normalising robotic avatar motion could be used to prevent further observation of the operator's emotions. The calibration of such a system and its impact on different telerobotic tasks would be valuable topics for future research.

Emotionally Intelligence Encounters with Robotic Avatars

By using similar emotion inference approaches, we could facilitate the recognition of naturalistic body language, trained on real human motion data, in both virtual and physical robotic avatars. VR allows people to be embodied within virtual environments and act within them using virtual avatars which can express their body language. Liebers *et al.* [281] found it is possible to identify individuals via their virtual body language in VR and it has been shown that virtual agents can express emotion through body language [282, 283]. Our methods could be applied to these virtual avatars, allowing for the automatic detection of users' emotions in VR settings. This could be used to tailor user experiences; for example, if during a VR game a user is expressing anger or sadness the game could dynamically become easier or calmer, as seen in prior work [284]. Furthermore, our approach could be leveraged to enable more emotionally intelligent interactions with the virtual world, NPCs and other users.

These emotionally intelligent encounters could also take place in real-world settings. In the wake of the COVID-19 pandemic, working from home has become more prominent as a current and future labour trend. In future, we may see physical robotic avatars, such as robotic arms, partially or fully replace human workplaces such as offices or factories. If these avatars could both express emotion and have their operator's emotion understood by other co-located humans

or robotic avatars, it would help maintain affective relationships commonplace in social and working contexts. Our work shows that robot arm avatars have inherent distinct movement traits from differing operator emotions. Humans already possess the ability to infer affect from human arm movement [138] and future work could now explore if this also extends to the emotive movements of robotic arms inherited from their operators.

Finally, future work could investigate how other robotic form factors may inherit emotions, such as quadruped robots, such as Boston Dynamic's *Spot*[1], more limited humanoid social robots such as *Pepper*[2] and*Sophia*[3], which can only articulate their heads, arms and torso, or robotic hands, such as *Shadow Hand*[4]. While telerobotic avatars inherent emotion from the operator's natural arm movements, these quadrupedal or social robots are instead operated using a controller, such as a gamepad, so whether emotion can be inferred from such control mechanisms should also be investigated. Finally, future full-bodied robotic avatars could inherit yet more complicated and nuanced emotional features, as more pronounced movement across the whole body is used to express emotive movement features and the relationship between different body parts can provide more emotional information, as has been explored with virtual agents [282].

4.9 Conclusion

This paper demonstrates that a motion-controlled robotic arm can inherit the human operator's emotions, then both describes and evaluates an approach for classifying human emotions based on motion-controlled robotic avatar motion behaviours in interactive control scenarios. We extracted the emotion-related features from robot end-effector data and developed a DTW-based algorithm to classify individual subjects' emotions. We further develop an alternative CNN-based algorithm to classify emotions. The training model used could be subject-dependent or independent. Analysis of a dataset of 6000 tasks using a motion-controlled robotic avatar platform found that our approach achieved up to 83.3% accuracy in recognizing the user's emotion. Our approach is highly suited to motion-based telerobotic use cases when compared to traditional methods. We discuss how this method can be applied to current remote robot operations to build efficient, safe and human-centred interactions. Furthermore, we explore promising future applications for this approach, including virtual robotic avatars, emotional intelligence encounters between man and machine and AI-assisted emotive-motion dampening.

[1]Spot by Boston Dynamics - https://bostondynamics.com/products/spot/ - Accessed 23/08/23.
[2]Pepper by Aldabaren Robotics - https://www.aldebaran.com/en/pepper - Accessed 23/08/23.
[3]Sophia by Hanson Robotics - https://www.hansonrobotics.com/sophia/ - Accessed 23/08/23.
[4]Shadow Hand by Shadow Robot - https://www.shadowrobot.com/ - Accessed 23/08/23.

Chapter 5

Conclusions and Future Works

This thesis classified users' identities and users' emotions using the motion-controlled robot's end-effector trajectory enhancing the teleoperation system security. In this thesis, we did extensive experiments and implemented machine learning and deep learning algorithms to analyse robotic data. We proposed a more holistic approach to studying motion-controlled robotic systems. This integrated perspective could significantly contribute to the fields of building secure, trustworthy, personalized and human-centered HRI.

Chapter 3 investigated user classification using motion-controlled robots. It showed that both motion-controlled Franka robot (7DoFs) and NAO (6DoF) robot can inherit human operators' biometric information in the hand trajectory, and they achieved comparable accuracy. Besides, we verified that the lower controlling system performance leads to lower user classification accuracy. Future work could design customized control systems based on users' personal control preferences. However, in some situations, user identities should be protected to secure users' privacy. Thus, we proposed a reinforcement learning method to protect user biometric information in robotic trajectory. We wiped human identity information in general, so future work could design more specific identity features wipe to adapt different users' control habits. User classification provided a foundation for emotion classification work. Different individuals have different ways of expressing emotions, so we performed emotion classification using a motion-controlled robotic arm's end-effector trajectory based on identified users in Chapter 4. Before this, Chapter 2.2 provided a comprehensive review of existing emotion recognition methods and their application for HRI. To our best knowledge, there is no work to study the emotion inherited by the motion-controlled robotic arm, but we can not underestimate the values. We selected one method that is most suitable for our motion-controlled robotic arm platform. We proposed methods of human emotion classification from a motion-controlled robotic arm's trajectory. In addition, we selected ECG, a common and well-established emotion recognition method, to verify our methods' efficiency. Our study focused on the basic emotions, however, future work could investigate emotions that highly appear in the teleoperated scenario. However, for both user classification and emotion classification experiment setup, we selected tasks

that mirror real-world activities, yet they were not conducted in actual real-world settings. Additionally, our focus was primarily on the use of robotic arms, while the future of teleoperated robots may encompass full-body control. Based on different robotic types and their related scenarios. Future research could explore a broader range of tasks in real-world Human-Robot Interaction (HRI) applications, considering various robotic types and their respective scenarios. For instance, future studies could develop manipulative motions tailored for industrial robots, involving the direct manipulation of objects such as picking up, moving, or assembling items. Additionally, there is potential to design locomotive motions, which involve the movement of the human body in space, such as walking, turning, and bending, that are useful in navigation or search and rescue operations applications. Moreover, future research could focus on designing sequential motions, which are complex tasks requiring a series of actions performed in a specific sequence, like assembling machine parts. These directions allow our proposed methods in diverse real-world scenarios, enhancing the applicability of teleoperated robots.

Bibliography

[1] MD Moniruzzaman, Alexander Rassau, Douglas Chai, and Syed Mohammed Shamsul Islam. Teleoperation methods and enhancement techniques for mobile robots: A comprehensive survey. *Robotics and Autonomous Systems*, 150:103973, 2022.

[2] Hocheol Shin, Seung Ho Jung, You Rack Choi, and ChangHoi Kim. Development of a shared remote control robot for aerial work in nuclear power plants. *Nuclear Engineering and Technology*, 50(4):613–618, 2018.

[3] Md Assad-Uz-Zaman, Md Rasedul Islam, Mohammad Habibur Rahman, Ying-Chih Wang, and Erin McGonigle. Kinect controlled nao robot for telerehabilitation. *Journal of Intelligent Systems*, 30(1):224–239, 2020.

[4] Tamás Haidegger, József Sándor, and Zoltán Benyó. Surgery in space: the future of robotic telesurgery. *Surgical Endoscopy*, 25:681–690, 2011.

[5] Benjamin Gleason and Christine Greenhow. Hybrid education: The potential of teaching and learning with robot-mediated communication. *Online Learning Journal*, 21(4), 2017.

[6] Jean-Paul A Yaacoub, Hassan N Noura, Ola Salman, and Ali Chehab. Robotics cyber security: Vulnerabilities, attacks, countermeasures, and recommendations. *International Journal of Information Security*, 21(1):115–158, 2022.

[7] Robert Morris and Ken Thompson. Password security: A case history. *Communications of the ACM*, 22(11):594–597, 1979.

[8] Alexander Chan, Tzipora Halevi, and Nasir Memon. Leap motion controller for authentication via hand geometry and gestures. In *2015 Third International Conference on Human Aspects of Information Security, Privacy, and Trust*, pages 13–22. Springer, 2015.

[9] Gerben A Van Kleef, Evert A Van Doorn, Marc W Heerdink, and Lukas F Koning. Emotion is for influence. *European Review of Social Psychology*, 22(1):114–163, 2011.

[10] Woong Yeol Joe and So Young Song. Applying human-robot interaction technology in retail industries. *International Journal of Mechanical Engineering and Robotics Research*, 8(6):839–844, 2019.

[11] Jeong-Sik Park, Ji-Hwan Kim, and Yung-Hwan Oh. Feature vector classification based speech emotion recognition for service robots. *IEEE Transactions on Consumer Electronics*, 55(3):1590–1596, 2009.

[12] David Martin, Martin O'neill, Susan Hubbard, and Adrian Palmer. The role of emotion in explaining consumer satisfaction and future behavioural intention. *Journal of Services Marketing*, 22(3):224–236, 2008.

[13] Chenli Lin, Yuanyuan Ren, and Aming Lu. The effectiveness of virtual reality games in improving cognition, mobility, and emotion in elderly post-stroke patients: A systematic review and meta-analysis. *Neurosurgical Review*, 46(1):167, 2023.

[14] Ann O'leary. Stress, emotion, and human immune function. *Psychological Bulletin*, 108(3):363, 1990.

[15] Soujanya Poria, Erik Cambria, Rajiv Bajpai, and Amir Hussain. A review of affective computing: From unimodal analysis to multimodal fusion. *Information Fusion*, 37:98–125, 2017.

[16] Stephanie Hui-Wen Chuah and Joanne Yu. The future of service: The power of emotion in human-robot interaction. *Journal of Retailing and Consumer Services*, 61:102551, 2021.

[17] Zhentao Liu, Min Wu, Weihua Cao, Luefeng Chen, Jianping Xu, Ri Zhang, Mengtian Zhou, and Junwei Mao. A facial expression emotion recognition based human-robot interaction system. *IEEE/CAA Journal of Automatica Sinica*, 4(4):668–676, 2017.

[18] Marco Leo, Marco Del Coco, Pierluigi Carcagni, Cosimo Distante, Massimo Bernava, Giovanni Pioggia, and Giuseppe Palestra. Automatic emotion recognition in robot-children interaction for asd treatment. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pages 145–153, 2015.

[19] Thomas B Sheridan. Human–robot interaction: status and challenges. *Human Factors*, 58(4):525–532, 2016.

[20] Jonathan Kofman, Xianghai Wu, Timothy J Luu, and Siddharth Verma. Teleoperation of a robot manipulator using a vision-based human-robot interface. *IEEE Transactions on Industrial Electronics*, 52(5):1206–1219, 2005.

[21] Geng Yang, Honghao Lv, Zhiyu Zhang, Liu Yang, Jia Deng, Siqi You, Juan Du, and Huayong Yang. Keep healthcare workers safe: application of teleoperated robot in isolation ward for covid-19 prevention and control. *Chinese Journal of Mechanical Engineering*, 33:1–4, 2020.

[22] Kiho Kim, Jangjin Park, Hohee Lee, and Keechan Song. Teleoperated cleaning robots for use in a highly radioactive environment of the dfdf. In *2006 SICE-ICASE International Joint Conference*, pages 3094–3099. IEEE, 2006.

[23] Bing Cai Kok and Harold Soh. Trust in robots: Challenges and opportunities. *Current Robotics Reports*, 1(4):297–309, 2020.

[24] Carsten Maple, Gregory Epiphaniou, Waleed Hathal, Ugur Ilker Atmaca, Haitham Cruickshank, Gregory Falco, et al. The impact of message encryption on teleoperation for space applications. In *2022 IEEE Aerospace Conference (AERO)*, pages 1–10. IEEE, 2022.

[25] Rajdeep Bhanot and Rahul Hans. A review and comparative analysis of various encryption algorithms. *International Journal of Security and Its Applications*, 9(4):289–306, 2015.

[26] Pinyao Guo, Hunmin Kim, Nurali Virani, Jun Xu, Minghui Zhu, and Peng Liu. Roboads: Anomaly detection against sensor and actuator misbehaviors in mobile robots. In *2018 48th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN)*, pages 574–585. IEEE, 2018.

[27] Jun Inoue, Yoriyuki Yamagata, Yuqi Chen, Christopher M Poskitt, and Jun Sun. Anomaly detection for a water treatment system using unsupervised machine learning. In *2017 IEEE International Conference on Data Mining Workshops (ICDMW)*, pages 1058–1065. IEEE, 2017.

[28] Syed W Shah and Salil S Kanhere. Recent trends in user authentication–a survey. *IEEE Access*, 7:112505–112519, 2019.

[29] Jing Tian, Chengzhang Qu, Wenyuan Xu, and Song Wang. Kinwrite: Handwriting-based authentication using kinect. In *NDSS*, volume 93, page 94, 2013.

[30] Grady Xiao, Mariofanna Milanova, and Mengjun Xie. Secure behavioral biometric authentication with leap motion. In *2016 4th International Symposium on Digital Forensic and Security (ISDFS)*, pages 112–118. IEEE, 2016.

[31] Junshuang Yang, Yanyan Li, and Mengjun Xie. Motionauth: Motion-based authentication for wrist worn smart devices. In *2015 IEEE International Conference on Pervasive Computing and Communication Workshops (PerCom Workshops)*, pages 550–555. IEEE, 2015.

[32] Aleksandr Ometov, Sergey Bezzateev, Niko Mäkitalo, Sergey Andreev, Tommi Mikkonen, and Yevgeni Koucheryavy. Multi-factor authentication: A survey. *Cryptography*, 2(1):1, 2018.

[33] Daniel Goleman. *Emotional intelligence*. Bloomsbury Publishing, 2020.

[34] Gabrielle Simcock, Larisa T McLoughlin, Tamara De Regt, Kathryn M Broadhouse, Denise Beaudequin, Jim Lagopoulos, and Daniel F Hermens. Associations between facial emotion recognition and mental health in early adolescence. *International Journal of Environmental Research and Public Health*, 17(1):330, 2020.

[35] Dongri Yang, Abeer Alsadoon, PW Chandana Prasad, Ashutosh Kumar Singh, and Amr Elchouemi. An emotion recognition model based on facial recognition in virtual learning environment. *Procedia Computer Science*, 125:2–10, 2018.

[36] Sonali T Saste and SM Jagdale. Emotion recognition from speech using mfcc and dwt for security system. In *2017 International Conference of Electronics, Communication and Aerospace Technology (ICECA)*, volume 1, pages 701–704. IEEE, 2017.

[37] Bernardete Ribeiro, Gonçalo Oliveira, Ana Laranjeira, and Joel P Arrais. Deep learning in digital marketing: brand detection and emotion recognition. *International Journal of Machine Intelligence and Sensory Signal Processing*, 2(1):32–50, 2017.

[38] Kai-Tai Song, Meng-Ju Han, and Shih-Chieh Wang. Speech signal-based emotion recognition and its application to entertainment robots. *Journal of the Chinese Institute of Engineers*, 37(1):14–25, 2014.

[39] Roddy Cowie, Ellen Douglas-Cowie, Nicolas Tsapatsoulis, George Votsis, Stefanos Kollias, Winfried Fellenz, and John G Taylor. Emotion recognition in human-computer interaction. *IEEE Signal processing magazine*, 18(1):32–80, 2001.

[40] Javier Marín-Morales, Carmen Llinares, Jaime Guixeres, and Mariano Alcañiz. Emotion recognition in immersive virtual reality: From statistics to affective computing. *Sensors*, 20(18):5163, 2020.

[41] Dhwani Mehta, Mohammad Faridul Haque Siddiqui, and Ahmad Y Javaid. Facial emotion recognition: A survey and real-world user experiences in mixed reality. *Sensors*, 18(2):416, 2018.

[42] SM Sarala, DH Sharath Yadav, and Asadullah Ansari. Emotionally adaptive driver voice alert system for advanced driver assistance system (adas) applications. In *2018 International Conference on Smart Systems and Inventive Technology (ICSSIT)*, pages 509–512. IEEE, 2018.

[43] Matteo Spezialetti, Giuseppe Placidi, and Silvia Rossi. Emotion recognition for human-robot interaction: Recent advances and future perspectives. *Frontiers in Robotics and AI*, 7:532279, 2020.

[44] Mohammad Reza Loghmani, Stefano Rovetta, and Gentiane Venture. Emotional intelligence in robots: Recognizing human emotions from daily-life gestures. In *2017 IEEE International Conference on Robotics and Automation (ICRA)*, pages 1677–1684. IEEE, 2017.

[45] Maja Pantic and Leon JM Rothkrantz. Toward an affect-sensitive multimodal human-computer interaction. *Proceedings of the IEEE*, 91(9):1370–1390, 2003.

[46] Michelle Karg, Ali-Akbar Samadani, Rob Gorbet, Kolja Kühnlenz, Jesse Hoey, and Dana Kulić. Body movements for affective expression: A survey of automatic recognition and generation. *IEEE Transactions on Affective Computing*, 4(4):341–359, 2013.

[47] Yiyuan Wang, Luke Hespanhol, and Martin Tomitsch. How can autonomous vehicles convey emotions to pedestrians? a review of emotionally expressive non-humanoid robots. *Multimodal Technologies and Interaction*, 5(12):84, 2021.

[48] Stanislav Ivanov, Faruk Seyitoğlu, and Martina Markova. Hotel managers' perceptions towards the use of robots: a mixed-methods approach. *Information Technology & Tourism*, 22:505–535, 2020.

[49] Yuqian Lu, Hao Zheng, Saahil Chand, Wanqing Xia, Zengkun Liu, Xun Xu, Lihui Wang, Zhaojun Qin, and Jinsong Bao. Outlook on human-centric manufacturing towards industry 5.0. *Journal of Manufacturing Systems*, 62:612–627, 2022.

[50] Rebecca Andreasson, Beatrice Alenljung, Erik Billing, and Robert Lowe. Affective touch in human–robot interaction: conveying emotion to the nao robot. *International Journal of Social Robotics*, 10(4):473–491, 2018.

[51] Nourhan Elfaramawy, Pablo Barros, German I Parisi, and Stefan Wermter. Emotion recognition from body expressions with a neural network architecture. In *Proceedings of the 5th International Conference on Human Agent Interaction*, pages 143–149, 2017.

[52] SJ Rosula Reyes, Keanu M Depano, Aaron Matthew A Velasco, John Chris T Kwong, and Carlos M Oppus. Face detection and recognition of the seven emotions via facial expression: Integration of machine learning algorithm into the nao robot. In *2020 5th International Conference on Control and Robotics Engineering (ICCRE)*, pages 25–29. IEEE, 2020.

[53] Pinar Uluer, Hatice Kose, Elif Gumuslu, and Duygun Erol Barkana. Experience with an affective robot assistant for children with hearing disabilities. *International Journal of Social Robotics*, pages 1–18, 2021.

[54] Hans-Jörg Vögel, Christian Süß, Thomas Hubregtsen, Viviane Ghaderi, Ronee Chadowitz, Elisabeth André, Nicholas Cummins, Björn Schuller, Jérôme Härri, Raphaël Troncy, et al. Emotion-awareness for intelligent vehicle assistants: A research agenda. In *Proceedings of the 1st International Workshop on Software Engineering for AI in Autonomous Systems*, pages 11–15, 2018.

[55] Hojjat Abdollahi, Mohammad H Mahoor, Rohola Zandie, Jarid Siewierski, and Sara H Qualls. Artificial emotional intelligence in socially assistive robots for older adults: a pilot study. *IEEE Transactions on Affective Computing*, 14(3):2020–2032, 2022.

[56] Martin Saerbeck and Christoph Bartneck. Perception of affect elicited by robot motion. In *2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 53–60. IEEE, 2010.

[57] Fatemeh Noroozi, Ciprian Adrian Corneanu, Dorota Kamińska, Tomasz Sapiński, Sergio Escalera, and Gholamreza Anbarjafari. Survey on emotional body gesture recognition. *IEEE transactions on affective computing*, 12(2):505–523, 2018.

[58] Byoung Chul Ko. A brief review of facial emotion recognition based on visual information. *Sensors*, 18(2):401, 2018.

[59] Zhihong Zeng, Maja Pantic, Glenn I Roisman, and Thomas S Huang. A survey of affect recognition methods: audio, visual and spontaneous expressions. In *Proceedings of the 9th International Conference on Multimodal Interfaces*, pages 126–133, 2007.

[60] Andrius Dzedzickis, Artūras Kaklauskas, and Vytautas Bucinskas. Human emotion recognition: Review of sensors and methods. *Sensors*, 20(3):592, 2020.

[61] Yan Wang, Wei Song, Wei Tao, Antonio Liotta, Dawei Yang, Xinlei Li, Shuyong Gao, Yixuan Sun, Weifeng Ge, Wei Zhang, et al. A systematic review on affective computing: Emotion models, databases, and recent advances. *Information Fusion*, 83:19–52, 2022.

[62] Rosalind W. Picard, Elias Vyzas, and Jennifer Healey. Toward machine emotional intelligence: Analysis of affective physiological state. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(10):1175–1191, 2001.

[63] Carroll E Izard. *Human emotions*. Springer Science & Business Media, 2013.

[64] Alexandra Weidemann and Nele Rußwinkel. The role of frustration in human–robot interaction–what is needed for a successful collaboration? *Frontiers in Psychology*, page 707, 2021.

[65] Anastasia Pampouchidou, Panagiotis G Simos, Kostas Marias, Fabrice Meriaudeau, Fan Yang, Matthew Pediaditis, and Manolis Tsiknakis. Automatic assessment of depression

based on visual cues: A systematic review. *IEEE Transactions on Affective Computing*, 10(4):445–470, 2017.

[66] Christoffer Holmgård, Georgios N Yannakakis, Karen-Inge Karstoft, and Henrik Steen Andersen. Stress detection for ptsd via the startlemart game. In *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction*, pages 523–528. IEEE, 2013.

[67] Giorgos Giannakakis, Dimitris Grigoriadis, Katerina Giannakaki, Olympia Simantiraki, Alexandros Roniotis, and Manolis Tsiknakis. Review on psychological stress detection using biosignals. *IEEE Transactions on Affective Computing*, 13(1):440–460, 2019.

[68] Muhammad Anas Hasnul, Nor Azlina Ab Aziz, Salem Alelyani, Mohamed Mohana, and Azlan Abd Aziz. Electrocardiogram-based emotion recognition systems and their applications in healthcare—a review. *Sensors*, 21(15):5015, 2021.

[69] Christos D Katsis, Nikolaos Katertsidis, George Ganiatsas, and Dimitrios I Fotiadis. Toward emotion recognition in car-racing drivers: A biosignal processing approach. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 38(3):502–512, 2008.

[70] Raul Fernandez and Rosalind W Picard. Modeling drivers' speech under stress. *Speech Communication*, 40(1-2):145–159, 2003.

[71] Oussama El Hammoumi, Fatimaezzahra Benmarrakchi, Nihal Ouherrou, Jamal El Kafi, and Ali El Hore. Emotion recognition in e-learning systems. In *2018 6th International Conference on Multimedia Computing and Systems (ICMCS)*, pages 1–6. IEEE, 2018.

[72] Agata Kołakowska, Agnieszka Landowska, Mariusz Szwoch, Wioleta Szwoch, and Michal R Wrobel. Emotion recognition and its applications. *Human-Computer Systems Interaction: Backgrounds and Applications 3*, pages 51–62, 2014.

[73] Jaiteg Singh, Gaurav Goyal, and Rupali Gill. Use of neurometrics to choose optimal advertisement method for omnichannel business. *Enterprise Information Systems*, 14(2):243–265, 2020.

[74] Klaus R Scherer et al. Psychological models of emotion. *The Neuropsychology of Emotion*, 137(3):137–162, 2000.

[75] Paul Ekman. An argument for basic emotions. *Cognition & Emotion*, 6(3-4):169–200, 1992.

[76] James A Russell and Albert Mehrabian. Evidence for a three-factor theory of emotions. *Journal of research in Personality*, 11(3):273–294, 1977.

[77] Robert Plutchik and Henry Kellerman. *Theories of emotion*, volume 1. Academic Press, 2013.

[78] Mingmin Zhao, Fadel Adib, and Dina Katabi. Emotion recognition using wireless signals. In *Proceedings of the 22nd Annual International Conference on Mobile Computing and Networking*, pages 95–108, 2016.

[79] Peter J Lang, Margaret M Bradley, Bruce N Cuthbert, et al. International affective picture system (iaps): Technical manual and affective ratings. *NIMH Center for the Study of Emotion and Attention*, 1(39-58):3, 1997.

[80] Sander Koelstra, Christian Muhl, Mohammad Soleymani, Jong-Seok Lee, Ashkan Yazdani, Touradj Ebrahimi, Thierry Pun, Anton Nijholt, and Ioannis Patras. Deap: A database for emotion analysis; using physiological signals. *IEEE Transactions on Affective Computing*, 3(1):18–31, 2011.

[81] Stamos Katsigiannis and Naeem Ramzan. Dreamer: A database for emotion recognition through eeg and ecg signals from wireless low-cost off-the-shelf devices. *IEEE Journal of Biomedical and Health Informatics*, 22(1):98–107, 2017.

[82] Carlos Busso, Murtaza Bulut, Chi-Chun Lee, Abe Kazemzadeh, Emily Mower, Samuel Kim, Jeannette N Chang, Sungbok Lee, and Shrikanth S Narayanan. Iemocap: Interactive emotional dyadic motion capture database. *Language Resources and Evaluation*, 42(4):335–359, 2008.

[83] Laurence Likforman-Sulem, Anna Esposito, Marcos Faundez-Zanuy, Stéphan Clémençon, and Gennaro Cordasco. Emothaw: A novel database for emotional state recognition from handwriting and drawing. *IEEE Transactions on Human-Machine Systems*, 47(2):273–284, 2017.

[84] Hatice Gunes and Massimo Piccardi. A bimodal face and body gesture database for automatic analysis of human nonverbal affective behavior. In *18th International Conference on Pattern Recognition (ICPR'06)*, volume 1, pages 1148–1153. IEEE, 2006.

[85] Mehmet Berkehan Akçay and Kaya Oğuz. Speech emotion recognition: Emotional models, databases, features, preprocessing methods, supporting modalities, and classifiers. *Speech Communication*, 116:56–76, 2020.

[86] Nourah Alswaidan and Mohamed El Bachir Menai. A survey of state-of-the-art approaches for emotion recognition in text. *Knowledge and Information Systems*, 62(8):2937–2987, 2020.

[87] Robert Horlings, Dragos Datcu, and Leon JM Rothkrantz. Emotion recognition using brain activity. In *Proceedings of the 9th International Conference on Computer Systems and Technologies and Workshop for PhD students in Computing*, pages II–1, 2008.

[88] Lin Shu, Jinyan Xie, Mingyue Yang, Ziyi Li, Zhenqi Li, Dan Liao, Xiangmin Xu, and Xinyi Yang. A review of emotion recognition using physiological signals. *Sensors*, 18(7):2074, 2018.

[89] Sharmeen M Saleem Abdullah Abdullah, Siddeeq Y Ameen Ameen, Mohammed AM Sadeeq, and Subhi Zeebaree. Multimodal emotion recognition using deep learning. *Journal of Applied Science and Technology Trends*, 2(02):52–58, 2021.

[90] Tony Belpaeme, James Kennedy, Aditi Ramachandran, Brian Scassellati, and Fumihide Tanaka. Social robots for education: A review. *Science Robotics*, 3(21):eaat5954, 2018.

[91] Junya Nakanishi, Itaru Kuramoto, Jun Baba, Kohei Ogawa, Yuichiro Yoshikawa, and Hiroshi Ishiguro. Continuous hospitality with social robots at a hotel. *SN Applied Sciences*, 2:1–13, 2020.

[92] Giuliana Ferrante, Gianpaolo Vitale, Amelia Licari, Laura Montalbano, Giovanni Pilato, Ignazio Infantino, Agnese Augello, and Stefania La Grutta. Social robots and therapeutic adherence: A new challenge in pediatric asthma? *Paediatric Respiratory Reviews*, 40:46–51, 2021.

[93] Hideki Kozima, Marek P Michalowski, and Cocoro Nakagawa. Keepon: A playful robot for research, therapy, and entertainment. *International Journal of Social Robotics*, 1:3–18, 2009.

[94] Daniele Giansanti. The social robot in rehabilitation and assistance: what is the future? In *Healthcare*, volume 9, page 244. MDPI, 2021.

[95] Building acceptance and trust in autonomous mobility, mar 2022.

[96] Sebastian Zepf, Javier Hernandez, Alexander Schmitt, Wolfgang Minker, and Rosalind W Picard. Driver emotion recognition for intelligent vehicles: A survey. *ACM Computing Surveys (CSUR)*, 53(3):1–30, 2020.

[97] Needs, wants and behaviour of 'drivers' and automated vehicle users today and into the future, mar 2022.

[98] Suaave. supporting acceptance of automated vehicle, mar 2022.

[99] Zhiyi Ma, Marwa Mahmoud, Peter Robinson, Eduardo Dias, and Lee Skrypchuk. Automatic detection of a driver's complex mental states. In *17th International Conference on Computational Science and Its Applications*, pages 678–691. Springer, 2017.

[100] Abhiram Kolli, Alireza Fasih, Fadi Al Machot, and Kyandoghere Kyamakya. Non-intrusive car driver's emotion recognition using thermal camera. In *Proceedings of the Joint INDS'11 & ISTET'11*, pages 1–5. IEEE, 2011.

[101] Andrés Ricardo Pérez-Riera, Raimundo Barbosa-Barros, Rodrigo Daminello-Raimundo, and Luiz Carlos de Abreu. Main artifacts in electrocardiography. *Annals of Noninvasive Electrocardiology*, 23(2):e12494, 2018.

[102] Ferdous Ahmed, ASM Hossain Bari, and Marina L Gavrilova. Emotion recognition from body movement. *IEEE Access*, 8:11761–11781, 2019.

[103] Mohamad A Eid and Hussein Al Osman. Affective haptics: Current research and future directions. *IEEE Access*, 4:26–40, 2015.

[104] Jonghwa Kim and Elisabeth André. Emotion recognition based on physiological changes in music listening. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(12):2067–2083, 2008.

[105] Yu-Liang Hsu, Jeen-Shing Wang, Wei-Chun Chiang, and Chien-Han Hung. Automatic ecg-based emotion recognition in music listening. *IEEE Transactions on Affective Computing*, 11(1):85–99, 2017.

[106] Pritam Sarkar and Ali Etemad. Self-supervised learning for ecg-based emotion recognition. In *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 3217–3221. IEEE, 2020.

[107] Pritam Sarkar and Ali Etemad. Self-supervised ecg representation learning for emotion recognition. *IEEE Transactions on Affective Computing*, 13(3):1541–1554, 2020.

[108] Terence KL Hui and R Simon Sherratt. Coverage of emotion recognition for common wearable biosensors. *Biosensors*, 8(2):30, 2018.

[109] Qiang Zhang, Xianxiang Chen, Qingyuan Zhan, Ting Yang, and Shanhong Xia. Respiration-based emotion recognition with deep learning. *Computers in Industry*, 92:84–90, 2017.

[110] Ahsan Noor Khan, Achintha Avin Ihalage, Yihan Ma, Baiyang Liu, Yujie Liu, and Yang Hao. Deep learning framework for subject-independent emotion detection using wireless signals. *Plos One*, 16(2):e0242946, 2021.

[111] Henning Metzmacher, Daniel Wölki, Carolin Schmidt, Jérôme Frisch, and Christoph van Treeck. Real-time human skin temperature analysis using thermal image recognition for thermal comfort assessment. *Energy and Buildings*, 158:1063–1078, 2018.

[112] Iris B Mauss and Michael D Robinson. Measures of emotion: A review. *Cognition and Emotion*, 23(2):209–237, 2009.

[113] Robert Jenke, Angelika Peer, and Martin Buss. Feature extraction and selection for emotion recognition from eeg. *IEEE Transactions on Affective computing*, 5(3):327–339, 2014.

[114] Prashant Lahane, Jay Jagtap, Aditya Inamdar, Nihal Karne, and Ritwik Dev. A review of recent trends in eeg based brain-computer interface. In *2019 International Conference on Computational Intelligence in Data Science (ICCIDS)*, pages 1–6. IEEE, 2019.

[115] Yisi Liu, Olga Sourina, and Minh Khoa Nguyen. Real-time eeg-based human emotion recognition and visualization. In *2010 International Conference on Cyberworlds*, pages 262–269. IEEE, 2010.

[116] Nazmi Sofian Suhaimi, James Mountstephens, Jason Teo, et al. Eeg-based emotion recognition: A state-of-the-art review of current trends and opportunities. *Computational Intelligence and Neuroscience*, 2020, 2020.

[117] Paul Ekman and Wallace V Friesen. Facial action coding system. *Environmental Psychology & Nonverbal Behavior*, 1978.

[118] Paweł Tarnowski, Marcin Kołodziej, Andrzej Majkowski, and Remigiusz J Rak. Emotion recognition using facial expressions. *Procedia Computer Science*, 108:1175–1184, 2017.

[119] Jiayu Shu, Mangtik Chiu, and Pan Hui. Emotion sensing for mobile computing. *IEEE Communications Magazine*, 57(11):84–90, 2019.

[120] Hatice Gunes and Massimo Piccardi. Automatic temporal segment detection and affect recognition from face and body display. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 39(1):64–84, 2008.

[121] Neha Jain, Shishir Kumar, Amit Kumar, Pourya Shamsolmoali, and Masoumeh Zareapoor. Hybrid deep neural networks for face emotion recognition. *Pattern Recognition Letters*, 115:101–106, 2018.

[122] Yousif Khaireddin and Zhuofa Chen. Facial emotion recognition: State of the art performance on fer2013. *arXiv preprint arXiv:2105.03588*, 2021.

[123] Shamoil Shaees, Hamad Naeem, Muhammad Arslan, Muhammad Rashid Naeem, Syed Hamza Ali, and Hamza Aldabbas. Facial emotion recognition using transfer learning. In *2020 International Conference on Computing and Information Technology (ICCIT-1441)*, pages 1–5. IEEE, 2020.

[124] Satyajit Nayak, Bingi Nagesh, Aurobinda Routray, and Monalisa Sarma. A human–computer interaction framework for emotion recognition through time-series thermal video sequences. *Computers & Electrical Engineering*, 93:107280, 2021.

[125] Patil Sangramand Awale R. N. Kehri Vikram, Ingle Rahul. Analysis of facial emg signal for emotion recognition using wavelet packet transform and svm. In *Machine Intelligence and Signal Analysis*, pages 247–257. Springer Singapore, 2019.

[126] Alexander Plopski, Teresa Hirzle, Nahal Norouzi, Long Qian, Gerd Bruder, and Tobias Langlotz. The eye in extended reality: A survey on gaze interaction and eye tracking in head-worn extended reality. *ACM Computing Surveys (CSUR)*, 55(3):1–39, 2022.

[127] Paweł Tarnowski, Marcin Kołodziej, Andrzej Majkowski, and Remigiusz Jan Rak. Eye-tracking analysis for emotion recognition. *Computational intelligence and neuroscience*, 2020, 2020.

[128] Jia Zheng Lim, James Mountstephens, and Jason Teo. Emotion recognition using eye-tracking: taxonomy, review and current challenges. *Sensors*, 20(8):2384, 2020.

[129] Ruhul Amin Khalil, Edward Jones, Mohammad Inayatullah Babar, Tariqullah Jan, Mohammad Haseeb Zafar, and Thamer Alhussain. Speech emotion recognition using deep learning techniques: A review. *IEEE Access*, 7:117327–117345, 2019.

[130] Zheng Lian, Bin Liu, and Jianhua Tao. Ctnet: Conversational transformer network for emotion recognition. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 29:985–1000, 2021.

[131] Bagus Tris Atmaja and Akira Sasou. Evaluating self-supervised speech representations for speech emotion recognition. *IEEE Access*, 10:124396–124407, 2022.

[132] Yong-Soo Seol, Dong-Joo Kim, and Han-Woo Kim. Emotion recognition from text using knowledge-based ann. In *ITC-CSCC: International Technical Conference on Circuits Systems, Computers and Communications*, pages 1569–1572, 2008.

[133] Shadi Shaheen, Wassim El-Hajj, Hazem Hajj, and Shady Elbassuoni. Emotion recognition from text based on automatically generated rules. In *2014 IEEE International Conference on Data Mining Workshop*, pages 383–392. IEEE, 2014.

[134] Francisca Adoma Acheampong, Henry Nunoo-Mensah, and Wenyu Chen. Transformer models for text-based emotion detection: a review of bert-based approaches. *Artificial Intelligence Review*, pages 1–41, 2021.

[135] SV Kedar, DS Bormane, Aaditi Dhadwal, Shiwali Alone, and Rashi Agarwal. Automatic emotion recognition through handwriting analysis: a review. In *2015 International Conference on Computing Communication Control and Automation*, pages 811–816. IEEE, 2015.

[136] Jiawen Han, George Chernyshov, Dingding Zheng, Peizhong Gao, Takuji Narumi, Katrin Wolf, and Kai Kunze. Sentiment pen: Recognizing emotional context based on handwriting features. In *Proceedings of the 10th Augmented Human International Conference 2019*, pages 1–8, 2019.

[137] Donald Glowinski, Nele Dael, Antonio Camurri, Gualtiero Volpe, Marcello Mortillaro, and Klaus Scherer. Toward a minimal representation of affective gestures. *IEEE Transactions on Affective Computing*, 2(2):106–118, 2011.

[138] Frank E Pollick, Helena M Paterson, Armin Bruderlin, and Anthony J Sanford. Perceiving affect from arm movement. *Cognition*, 82(2):B51–B61, 2001.

[139] Danilo Avola, Luigi Cinque, Alessio Fagioli, Gian Luca Foresti, and Cristiano Massaroni. Deep temporal analysis for non-acted body affect recognition. *IEEE Transactions on Affective Computing*, 13(3):1366–1377, 2020.

[140] Tomasz Sapiński, Dorota Kamińska, Adam Pelikant, and Gholamreza Anbarjafari. Emotion recognition from skeletal movements. *Entropy*, 21(7):646, 2019.

[141] Simon Senecal, Louis Cuel, Andreas Aristidou, and Nadia Magnenat-Thalmann. Continuous body emotion recognition system during theater performances. *Computer Animation and Virtual Worlds*, 27(3-4):311–320, 2016.

[142] Daniel Bernhardt and Peter Robinson. Detecting affect from non-stylised body motions. In *International Conference on Affective Computing and Intelligent Interaction*, pages 59–70. Springer, 2007.

[143] Gentiane Venture, Hideki Kadone, Tianxiang Zhang, Julie Grèzes, Alain Berthoz, and Halim Hicheur. Recognizing emotions conveyed by human gait. *International Journal of Social Robotics*, 6(4):621–632, 2014.

[144] Yajurv Bhatia, ASM Hossain Bari, Gee-Sern Jison Hsu, and Marina Gavrilova. Motion capture sensor-based emotion recognition using a bi-modular sequential neural network. *Sensors*, 22(1):403, 2022.

[145] Ing Ding Jr and Meng-Chuan Hsieh. A hand gesture action-based emotion recognition system by 3d image sensor information derived from leap motion sensors for the specific group with restlessness emotion problems. *Microsystem Technologies*, pages 1–13, 2020.

[146] Liam Schoneveld, Alice Othmani, and Hazem Abdelkawy. Leveraging recent advances in deep learning for audio-visual emotion recognition. *Pattern Recognition Letters*, 146:1–7, 2021.

[147] Siyu Zhu, Jin Qi, Jie Hu, and Sheng Hao. A new approach for product evaluation based on integration of eeg and eye-tracking. *Advanced Engineering Informatics*, 52:101601, 2022.

[148] Loic Kessous, Ginevra Castellano, and George Caridakis. Multimodal emotion recognition in speech-based interaction using facial expression, body gesture and acoustic analysis. *Journal on Multimodal User Interfaces*, 3(1):33–48, 2010.

[149] Shamane Siriwardhana, Tharindu Kaluarachchi, Mark Billinghurst, and Suranga Nanayakkara. Multimodal emotion recognition with transformer-based self supervised feature fusion. *IEEE Access*, 8:176274–176285, 2020.

[150] TS Ashwin and Ram Mohana Reddy Guddeti. Affective database for e-learning and classroom environments using indian students' faces, hand gestures and body postures. *Future Generation Computer Systems*, 108:334–348, 2020.

[151] Arne Öhman, Alfons Hamm, and Kenneth Hugdahl. Cognition and the autonomic nervous system: orienting, anticipation, and conditioning. 2000.

[152] Robert W Levenson. The autonomic nervous system and emotion. *Emotion Review*, 6(2):100–112, 2014.

[153] Rollin McCraty. *Science of the heart: Exploring the role of the heart in human performance*. HeartMath Research Center, Institute of HeartMath, 2015.

[154] Foteini Agrafioti, Dimitris Hatzinakos, and Adam K Anderson. Ecg pattern analysis for emotion detection. *IEEE Transactions on Affective Computing*, 3(1):102–115, 2011.

[155] Hisanori Kataoka, Hiroshi Kano, Hiroaki Yoshida, Atsuo Saijo, Masashi Yasuda, and Masato Osumi. Development of a skin temperature measuring system for non-contact stress evaluation. In *Proceedings of the 20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society.*, volume 2, pages 940–943. IEEE, 1998.

[156] Muhammad Adeel Mahmood, Winston KG Seah, and Ian Welch. Reliability in wireless sensor networks: A survey and challenges ahead. *Computer Networks*, 79:166–187, 2015.

[157] Jianhua Zhang, Zhong Yin, Peng Chen, and Stefano Nichele. Emotion recognition using multi-modal data and machine learning techniques: A tutorial and review. *Information Fusion*, 59:103–126, 2020.

[158] Peixiang Zhong, Di Wang, and Chunyan Miao. Eeg-based emotion recognition using regularized graph neural networks. *IEEE Transactions on Affective Computing*, 13(3):1290–1301, 2020.

[159] Dasa Gorjan, Klaus Gramann, Kevin De Pauw, and Uros Marusic. Removal of movement-induced eeg artifacts: current state of the art and guidelines. *Journal of Neural Engineering*, 19(1):011004, 2022.

[160] Felipe Zago Canal, Tobias Rossi Müller, Jhennifer Cristine Matias, Gustavo Gino Scotton, Antonio Reis de Sa Junior, Eliane Pozzebon, and Antonio Carlos Sobieranski. A survey on facial emotion recognition techniques: A state-of-the-art literature review. *Information Sciences*, 582:593–617, 2022.

[161] Shan Li and Weihong Deng. Deep facial expression recognition: A survey. *IEEE Transactions on Affective Computing*, 13(3):1195–1215, 2020.

[162] M Rosario González-Rodríguez, M Carmen Díaz-Fernández, and Carmen Pacheco Gómez. Facial-expression recognition: An emergent approach to the measurement of tourist satisfaction through emotions. *Telematics and Informatics*, 51:101404, 2020.

[163] Pradeep Buddharaju, Ioannis T Pavlidis, Panagiotis Tsiamyrtzis, and Mike Bazakos. Physiology-based face recognition in the thermal infrared spectrum. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(4):613–626, 2007.

[164] Jingu Heo, Seong G Kong, Besma R Abidi, and Mongi A Abidi. Fusion of visual and thermal signatures with eyeglass removal for robust face recognition. In *2004 Conference on Computer Vision and Pattern Recognition Workshop*, pages 122–122. IEEE, 2004.

[165] Diego A Socolinsky, Andrea Selinger, and Joshua D Neuheisel. Face recognition with visible and thermal infrared imagery. *Computer Vision and Image Understanding*, 91(1-2):72–114, 2003.

[166] Chiara Filippini, David Perpetuini, Daniela Cardone, Antonio Maria Chiarelli, and Arcangelo Merla. Thermal infrared imaging-based affective computing and its application to facilitate human robot interaction: A review. *Applied Sciences*, 10(8):2924, 2020.

[167] Wataru Sato, Koichi Murata, Yasuyuki Uraoka, Kazuaki Shibata, Sakiko Yoshikawa, and Masafumi Furuta. Emotional valence sensing using a wearable facial emg device. *Scientific Reports*, 11(1):5757, 2021.

[168] Shraddha A Mithbavkar and Milind S Shah. Analysis of emg based emotion recognition for multiple people and emotions. In *2021 IEEE 3rd Eurasia Conference on Biomedical Engineering, Healthcare and Sustainability (ECBIOS)*, pages 1–4. IEEE, 2021.

[169] Robert JK Jacob and Keith S Karn. Eye tracking in human-computer interaction and usability research: Ready to deliver the promises. In *The Mind's Eye*, pages 573–605. Elsevier, 2003.

[170] Melissa H Black, Nigel TM Chen, Kartik K Iyer, Ottmar V Lipp, Sven Bölte, Marita Falkmer, Tele Tan, and Sonya Girdler. Mechanisms of facial emotion recognition in autism spectrum disorders: Insights from eye tracking and electroencephalography. *Neuroscience & Biobehavioral Reviews*, 80:488–515, 2017.

[171] Wei-Long Zheng, Bo-Nan Dong, and Bao-Liang Lu. Multimodal emotion recognition using eeg and eye tracking data. In *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 5040–5043. IEEE, 2014.

[172] Egor Lakomkin, Mohammad Ali Zamani, Cornelius Weber, Sven Magg, and Stefan Wermter. On the robustness of speech emotion recognition for human-robot interaction with deep neural networks. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 854–860. IEEE, 2018.

[173] Daniela Cardenas. Handwriting analysis (graphology). *Wr/Rd093*.

[174] Samira Ebrahimi Kahou, Xavier Bouthillier, Pascal Lamblin, Caglar Gulcehre, Vincent Michalski, Kishore Konda, Sébastien Jean, Pierre Froumenty, Yann Dauphin, Nicolas Boulanger-Lewandowski, et al. Emonets: Multimodal deep learning approaches for emotion recognition in video. *Journal on Multimodal User Interfaces*, 10(2):99–111, 2016.

[175] Dongmin Shin, Dongil Shin, and Dongkyoo Shin. Development of emotion recognition interface using complex eeg/ecg bio-signal for interactive contents. *Multimedia Tools and Applications*, 76:11449–11470, 2017.

[176] Jinting Wu, Yujia Zhang, Shiying Sun, Qianzhong Li, and Xiaoguang Zhao. Generalized zero-shot emotion recognition from body gestures. *Applied Intelligence*, pages 1–19, 2022.

[177] Diego Resende Faria, Mario Vieira, and Fernanda CC Faria. Towards the development of affective facial expression recognition for human-robot interaction. In *Proceedings of the 10th International Conference on Pervasive Technologies Related to Assistive Environments*, pages 300–304, 2017.

[178] Mingyang Shao, Silas Franco Dos Reis Alves, Omar Ismail, Xinyi Zhang, Goldie Nejat, and Beno Benhabib. You are doing great! only one rep left: an affect-aware social robot for exercising. In *2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)*, pages 3811–3817. IEEE, 2019.

[179] Andrius Dzedzickis, Jurga Subačiūtė-Žemaitienė, Ernestas Šutinys, Urtė Samukaitė-Bubnienė, and Vytautas Bučinskas. Advanced applications of industrial robotics: New trends and possibilities. *Applied Sciences*, 12(1):135, 2021.

[180] Andrei Cotruş, Camelia Stanciu, and Alina Andreea Bulborea. Eq vs. iq which is most important in the success or failure of a student? *Procedia-Social and Behavioral Sciences*, 46:5211–5213, 2012.

[181] Tehao Zhu, Zeyang Xia, Jiaqi Dong, and Qunfei Zhao. A sociable human-robot interaction scheme based on body emotion analysis. *International Journal of Control, Automation and Systems*, 17(2):474–485, 2019.

[182] Seyed Farokh Atashzar, Michael Naish, and Rajni V Patel. Active sensorimotor augmentation in robotics-assisted surgical systems. In *Mixed and Augmented Reality in Medicine*, pages 61–81. CRC Press, 2018.

[183] Sandra Hirche and Martin Buss. Human-oriented control for haptic teleoperation. *Proceedings of the IEEE*, 100(3):623–647, 2012.

[184] Zheng Chen, Fanghao Huang, Weichao Sun, Jason Gu, and Bin Yao. Rbf-neural-network-based adaptive robust control for nonlinear bilateral teleoperation manipulators with uncertainty and time delay. *Ieee/Asme Transactions on Mechatronics*, 25(2):906–918, 2019.

[185] Yuheng Fan, Chenguang Yang, and Xinyu Wu. Improved teleoperation of an industrial robot arm system using leap motion and myo armband. In *2019 IEEE International Conference on Robotics and Biomimetics (ROBIO)*, pages 1670–1675. IEEE, 2019.

[186] Xin Wu, Canjun Yang, Yuanchao Zhu, Weitao Wu, and Qianxiao Wei. An integrated vision-based system for efficient robot arm teleoperation. *Industrial Robot: The international Journal of Robotics Research and Application*, 48(2):199–210, 2021.

[187] Ziwei Wang, Zhang Chen, and Bin Liang. Fixed-time velocity reconstruction scheme for space teleoperation systems: Exp barrier lyapunov function approach. *Acta Astronautica*, 157:92–101, 2019.

[188] Andreas Birk, Sören Schwertfeger, and Kaustubh Pathak. A networking framework for teleoperation in safety, security, and rescue robotics. *IEEE Wireless Communications*, 16(1):6–13, 2009.

[189] Tomáš Kot and Petr Novák. Application of virtual reality in teleoperation of the military mobile robotic system taros. *International Journal of Advanced Robotic Systems*, 15(1):1729881417751545, 2018.

[190] Sarmad Mehrdad, Fei Liu, Minh Tu Pham, Arnaud Lelevé, and S Farokh Atashzar. Review of advanced medical telerobots. *Applied Sciences*, 11(1):209, 2020.

[191] Mauricio Marcano, Sergio Díaz, Joshué Pérez, and Eloy Irigoyen. A review of shared control for automated vehicles: Theory and applications. *IEEE Transactions on Human-Machine Systems*, 50(6):475–491, 2020.

[192] Jiewu Leng, Weinan Sha, Baicun Wang, Pai Zheng, Cunbo Zhuang, Qiang Liu, Thorsten Wuest, Dimitris Mourtzis, and Lihui Wang. Industry 5.0: Prospect and retrospect. *Journal of Manufacturing Systems*, 65:279–295, 2022.

[193] Abdelfetah Hentout, Mustapha Aouache, Abderraouf Maoudj, and Isma Akli. Human–robot interaction in industrial collaborative robotics: a literature review of the decade 2008–2017. *Advanced Robotics*, 33(15-16):764–799, 2019.

[194] Maria Egger, Matthias Ley, and Sten Hanke. Emotion recognition from physiological signal analysis: A review. *Electronic Notes in Theoretical Computer Science*, 343:35–55, 2019.

[195] Toshihiro Takeshita, Manabu Yoshida, Yusuke Takei, Atsushi Ouchi, Akinari Hinoki, Hiroo Uchida, and Takeshi Kobayashi. Relationship between contact pressure and motion artifacts in ecg measurement with electrostatic flocked electrodes fabricated on textile. *Scientific Reports*, 9(1):5897, 2019.

[196] DA Tong, KA Bartels, and KS Honeyager. Adaptive reduction of motion artifact in the electrocardiogram. In *Proceedings of the Second Joint 24th Annual Conference and the Annual Fall Meeting of the Biomedical Engineering Society][Engineering in Medicine and Biology*, volume 2, pages 1403–1404. IEEE, 2002.

[197] Dongyeol Seok, Sanghyun Lee, Minjae Kim, Jaeouk Cho, and Chul Kim. Motion artifact removal techniques for wearable eeg and ppg sensor systems. *Frontiers in Electronics*, 2:685513, 2021.

[198] Serdar Kucuk and Zafer Bingul. *Robot kinematics: Forward and inverse kinematics*. INTECH Open Access Publisher London, UK, 2006.

[199] Richard P Paul. *Robot manipulators: mathematics, programming, and control: the computer control of robot manipulators*. Richard Paul, 1981.

[200] Laith Alzubaidi, Jinglan Zhang, Amjad J Humaidi, Ayad Al-Dujaili, Ye Duan, Omran Al-Shamma, José Santamaría, Mohammed A Fadhel, Muthana Al-Amidie, and Laith Farhan. Review of deep learning: Concepts, cnn architectures, challenges, applications, future directions. *Journal of Big Data*, 8:1–74, 2021.

[201] Samer Hijazi, Rishi Kumar, Chris Rowen, et al. Using convolutional neural networks for image recognition. *Cadence Design Systems Inc.: San Jose, CA, USA*, 9(1), 2015.

[202] Wenpeng Yin, Katharina Kann, Mo Yu, and Hinrich Schütze. Comparative study of cnn and rnn for natural language processing. *arXiv preprint arXiv:1702.01923*, 2017.

[203] Dimitri Palaz, Ronan Collobert, et al. Analysis of cnn-based speech recognition system using raw speech as input. Technical report, Idiap, 2015.

[204] Gustavo A Casan, Enric Cervera, Amine A Moughlbay, Jaime Alemany, and Philippe Martinet. Ros-based online robot programming for remote education and training. In *2015 IEEE International Conference on Robotics and Automation (ICRA)*, pages 6101–6106. IEEE, 2015.

[205] Michael J Sobrepera, Vera G Lee, Suveer Garg, Rochelle Mendonca, and Michelle J Johnson. Perceived usefulness of a social robot augmented telehealth platform by therapists in the united states. *IEEE robotics and automation letters*, 6(2):2946–2953, 2021.

[206] Tom CT van Riet, Kevin TH Chin Jen Sem, Jean-Pierre TF Ho, René Spijker, Jens Kober, and Jan de Lange. Robot technology in dentistry, part one of a systematic review: literature characteristics. *Dental Materials*, 37(8):1217–1226, 2021.

[207] Marissa D'Souza, Julian Gendreau, Austin Feng, Lily H Kim, Allen L Ho, and Anand Veeravagu. Robotic-assisted spine surgery: history, efficacy, cost, and future trends. *Robotic Surgery: Research and Reviews*, pages 9–23, 2019.

[208] Franka official website, mar 2022.

[209] Optitrack official website, mar 2022.

[210] Pichao Wang, Zhaoyang Li, Yonghong Hou, and Wanqing Li. Action recognition based on joint trajectory maps using convolutional neural networks. In *Proceedings of the 24th ACM International Conference on Multimedia*, pages 102–106, 2016.

[211] Giacomo Giorgi, Fabio Martinelli, Andrea Saracino, and Mina Sheikhalishahi. Walking through the deep: Gait analysis for user authentication through deep learning. In *33rd IFIP TC 11 International Conference on ICT Systems Security and Privacy Protection*, pages 62–76. Springer, 2018.

[212] Christopher G Atkeson and John M Hollerbach. Kinematic features of unrestrained vertical arm movements. *Journal of Neuroscience*, 5(9):2318–2330, 1985.

[213] Penny Chong, Yuval Elovici, and Alexander Binder. User authentication based on mouse dynamics using deep neural networks: A comprehensive study. *IEEE Transactions on Information Forensics and Security*, 15:1086–1101, 2019.

[214] Ricardo M Marcacini, Julio C Carnevali, and João Domingos. On combining websensors and dtw distance for knn time series forecasting. In *2016 23rd International Conference on Pattern Recognition (ICPR)*, pages 2521–2525. IEEE, 2016.

[215] Daniel Rakita, Bilge Mutlu, and Michael Gleicher. Effects of onset latency and robot speed delays on mimicry-control teleoperation. In *HRI'20: Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, 2020.

[216] Donald Knuth. Reinforcement learning.

[217] Neziha Akalin and Amy Loutfi. Reinforcement learning approaches in social robotics. *Sensors*, 21(4):1292, 2021.

[218] Gary Sharp, Lorna Bourke, and Matthew JFX Rickard. Review of emotional intelligence in health care: an introduction to emotional intelligence for surgeons. *ANZ Journal of Surgery*, 90(4):433–440, 2020.

[219] Sandeep K Nayar, Liam Musto, Gautom Baruah, Roland Fernandes, and Rasiah Bharathan. Self-assessment of surgical skills: a systematic review. *Journal of Surgical Education*, 77(2):348–361, 2020.

[220] Jiayu Shu, Mangtik Chiu, and Pan Hui. Emotion sensing for mobile computing. *IEEE Communications Magazine*, 57(11):84–90, 2019.

[221] Aly Khalifa, Ahmed A Abdelrahman, Dominykas Strazdas, Jan Hintz, Thorsten Hempel, and Ayoub Al-Hamadi. Face recognition and tracking framework for human–robot interaction. *Applied Sciences*, 12(11):5568, 2022.

[222] Qi-rong Mao, Xin-yu Pan, Yong-zhao Zhan, and Xiang-jun Shen. Using kinect for real-time emotion recognition via facial expressions. *Frontiers of Information Technology & Electronic Engineering*, 16(4):272–282, 2015.

[223] Fatemeh Noroozi, Ciprian Adrian Corneanu, Dorota Kamińska, Tomasz Sapiński, Sergio Escalera, and Gholamreza Anbarjafari. Survey on emotional body gesture recognition. *IEEE Transactions on Affective Computing*, 12(2):505–523, 2021.

[224] Long Huang, Zhen Meng, Zeyu Deng, Chen Wang, Liying Li, and Guodong Zhao. Towards verifying the user of motion-controlled robotic arm systems via the robot behavior. *IEEE Internet of Things Journal*, 2021.

[225] Long Huang, Zhen Meng, Zeyu Deng, Chen Wang, Liying Li, and Guodong Zhao. Robot behavior-based user authentication for motion-controlled robotic systems. In *IEEE INFOCOM 2021-IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, pages 1–6. IEEE, 2021.

[226] Seyedeh Maryam Fakhrhosseini and Myounghoon Jeon. Affect/emotion induction methods. In *Emotions and Affect in Human Factors and Human-Computer Interaction*, pages 235–253. Elsevier, 2017.

[227] James A Russell. A circumplex model of affect. *Journal of Personality and Social Psychology*, 39(6):1161, 1980.

[228] Mingmin Zhao, Fadel Adib, and Dina Katabi. Emotion recognition using wireless signals. In *Proceedings of the 22nd Annual International Conference on Mobile Computing and Networking*, pages 95–108, 2016.

[229] RS Soundariya and R Renuga. Eye movement based emotion recognition using electrooculography. In *2017 Innovations in Power and Advanced Computing Technologies (i-PACT)*, pages 1–5. IEEE, 2017.

[230] Wei Xue and Tao Li. Aspect based sentiment analysis with gated convolutional networks. *arXiv preprint arXiv:1805.07043*, 2018.

[231] Jianhai Zhang, Ming Chen, Shaokai Zhao, Sanqing Hu, Zhiguo Shi, and Yu Cao. Relieff-based eeg sensor selection methods for emotion recognition. *Sensors*, 16(10):1558, 2016.

[232] Theekshana Dissanayake, Yasitha Rajapaksha, Roshan Ragel, and Isuru Nawinne. An ensemble learning approach for electrocardiogram sensor based human emotion recognition. *Sensors*, 19(20):4495, 2019.

[233] M. Pantic and L.J.M. Rothkrantz. Toward an affect-sensitive multimodal human-computer interaction. *Proceedings of the IEEE*, 91(9):1370–1390, 2003.

[234] Michelle Karg, Ali-Akbar Samadani, Rob Gorbet, Kolja Kühnlenz, Jesse Hoey, and Dana Kulić. Body movements for affective expression: A survey of automatic recognition and generation. *IEEE Transactions on Affective Computing*, 4(4):341–359, 2013.

[235] Asha Kapur, Ajay Kapur, Naznin Virji-Babul, George Tzanetakis, and Peter F Driessen. Gesture-based affective computing on motion capture data. In *Affective Computing and Intelligent Interaction: First International Conference, ACII, 2005. Proceedings 1*, pages 1–7. Springer, 2005.

[236] Marko Lugger and Bin Yang. Cascaded emotion classification via psychological emotion dimensions using a large set of voice quality parameters. In *2008 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 4945–4948. IEEE, 2008.

[237] Donald Glowinski, Nele Dael, Antonio Camurri, Gualtiero Volpe, Marcello Mortillaro, and Klaus Scherer. Toward a minimal representation of affective gestures. *IEEE Transactions on Affective Computing*, 2(2):106–118, 2011.

[238] Harald G Wallbott. Bodily expression of emotion. *European Journal of Social Psychology*, 28(6):879–896, 1998.

[239] Daniel Bernhardt. Emotion inference from human body motion. Technical report, University of Cambridge, Computer Laboratory, 2010.

[240] Yann Maret, Daniel Oberson, and Marina Gavrilova. Identifying an emotional state from body movements using genetic-based algorithms. In *17th International Conference on Artificial Intelligence and Soft Computing (ICAISC)*, pages 474–485. Springer, 2018.

[241] Mei Silviana Saputri, Rahmad Mahendra, and Mirna Adriani. Emotion classification on indonesian twitter dataset. In *2018 International Conference on Asian Language Processing (IALP)*, pages 90–95, 2018.

[242] Liying Yang and Sheng-Feng Qin. A review of emotion recognition methods from keystroke, mouse, and touchscreen dynamics. *IEEE Access*, 9:162197–162213, 2021.

[243] Agata Kołakowska. A review of emotion recognition methods based on keystroke dynamics and mouse movements. In *2013 6th International Conference on Human System Interactions (HSI)*, pages 548–555. IEEE, 2013.

[244] Moojan Ghafurian, Gabriella Lakatos, and Kerstin Dautenhahn. The Zoomorphic Miro Robot's Affective Expression Design and Perceived Appearance. *International Journal of Social Robotics*, 14(4):945–962, 2022.

[245] Junchao Xu, Joost Broekens, Koen Hindriks, and Mark A Neerincx. *Mood expression through parameterized functional behavior of robots*. IEEE, 2013.

[246] Roshni Kaushik and Reid Simmons. Perception of emotion in torso and arm movements on humanoid robot quori. In *Companion of the 2021 ACM/IEEE International Conference on Human-Robot Interaction*, HRI '21 Companion, page 62–66. Association for Computing Machinery, 2021.

[247] Ros official website, mar 2022.

[248] Universal robots official website, mar 2022.

[249] Moveit official website, mar 2022.

[250] R. Cowie, E. Douglas-Cowie, N. Tsapatsoulis, G. Votsis, S. Kollias, W. Fellenz, and J.G. Taylor. Emotion recognition in human-computer interaction. *IEEE Signal Processing Magazine*, 18(1):32–80, 2001.

[251] Zheng Zhang, Jeff M Girard, Yue Wu, Xing Zhang, Peng Liu, Umur Ciftci, Shaun Canavan, Michael Reale, Andy Horowitz, Huiyuan Yang, et al. Multimodal spontaneous emotion corpus for human behavior analysis. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3438–3446, 2016.

[252] John G Taylor and Nickolaos F Fragopanagos. The interaction of attention and emotion. *Neural Networks*, 18(4):353–369, 2005.

[253] Jonathan Mitchell. Emotion and attention. *Philosophical Studies*, 180(1):73–99, 2023.

[254] James A Russell and Albert Mehrabian. Evidence for a three-factor theory of emotions. *Journal of Research in Personality*, 11(3):273–294, 1977.

[255] Shaun Alexander Macdonald, Frank Pollick, and Stephen Anthony Brewster. The impact of thermal cues on affective responses to emotionally resonant vibrations. In *Proceedings of the 2022 International Conference on Multimodal Interaction*, pages 259–269, 2022.

[256] Sidney D'Mello and Rafael A Calvo. Beyond the basic emotions: what should affective computing compute? In *CHI'13 Extended Abstracts on Human Factors in Computing Systems*, pages 2287–2294. 2013.

[257] Nigel Bosch, Sidney D'Mello, Ryan Baker, Jaclyn Ocumpaugh, Valerie Shute, Matthew Ventura, Lubin Wang, and Weinan Zhao. Automatic detection of learning-centered affective states in the wild. In *Proceedings of the 20th International Conference on Intelligent User Interfaces*, pages 379–388, 2015.

[258] Laurence Likforman-Sulem, Anna Esposito, Marcos Faundez-Zanuy, Stéphan Clémençon, and Gennaro Cordasco. Emothaw: A novel database for emotional state recognition from handwriting and drawing. *IEEE Transactions on Human-Machine Systems*, 47(2):273–284, 2017.

[259] Malika Arora, Munish Kumar, and Naresh Kumar Garg. Facial emotion recognition system based on pca and gradient features. *National Academy Science Letters*, 41:365–368, 2018.

[260] Diego Felipe Paez Granados, Breno A Yamamoto, Hiroko Kamide, Jun Kinugawa, and Kazuhiro Kosuge. Dance teaching by a robot: Combining cognitive and physical human–robot interaction for supporting the skill learning process. *IEEE Robotics and Automation Letters*, 2(3):1452–1459, 2017.

[261] Emily L Hampp, Morad Chughtai, Laura Y Scholl, Nipun Sodhi, Manoshi Bhowmik-Stoker, David J Jacofsky, and Michael A Mont. Robotic-arm assisted total knee arthroplasty demonstrated greater accuracy and precision to plan compared with manual techniques. *The Journal of Knee Surgery*, 32(03):239–250, 2019.

[262] Yiqun Wu, Feng Wang, Shengchi Fan, and James Kwok-Fai Chow. Robotics in dental implantology. *Oral and Maxillofacial Surgery Clinics*, 31(3):513–518, 2019.

[263] Jim Mainprice, Rafi Hayne, and Dmitry Berenson. Predicting human reaching motion in collaborative tasks using inverse optimal control and iterative re-planning. In *2015 IEEE International Conference on Robotics and Automation (ICRA)*, pages 885–892. IEEE, 2015.

[264] Claudia Pérez-D'Arpino and Julie A Shah. Fast target prediction of human reaching motion for cooperative human-robot manipulation tasks using time series classification. In *2015 IEEE International Conference on Robotics and Automation (ICRA)*, pages 6175–6182. IEEE, 2015.

[265] Neska El Haouij, Jean-Michel Poggi, Raja Ghozi, Sylvie Sevestre-Ghalila, and Mériem Jaïdane. Random forest-based approach for physiological functional variable selection for driver's stress level classification. *Statistical Methods & Applications*, 28:157–185, 2019.

[266] Ahmet Akbas. Evaluation of the physiological data indicating the dynamic stress level of drivers. *Scientific Research and Essays*, 6(2):430–439, 2011.

[267] N Keshan, PV Parimi, and Isabelle Bichindaritz. Machine learning for stress detection from ecg signals in automobile drivers. In *2015 IEEE International Conference on Big Data*, pages 2661–2669. IEEE, 2015.

[268] Nermine Munla, Mohamad Khalil, Ahmad Shahin, and Azzam Mourad. Driver stress level detection using hrv analysis. In *2015 International Conference on Advances in Biomedical Engineering (ICABME)*, pages 61–64. IEEE, 2015.

[269] Wen Wu, Chao Zhang, Xixin Wu, and Philip C Woodland. Estimating the uncertainty in emotion class labels with utterance-specific dirichlet priors. *IEEE Transactions on Affective Computing*, 2022.

[270] Javier Marín-Morales, Juan Luis Higuera-Trujillo, Alberto Greco, Jaime Guixeres, Carmen Llinares, Enzo Pasquale Scilingo, Mariano Alcañiz, and Gaetano Valenza. Affective computing in virtual reality: emotion recognition from brain and heartbeat dynamics using wearable sensors. *Scientific Reports*, 8(1):13657, 2018.

[271] Lisa Torrey and Jude Shavlik. Transfer learning. In *Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods, and Techniques*, pages 242–264. IGI global, 2010.

[272] Yuan-Pin Lin and Tzyy-Ping Jung. Improving eeg-based emotion classification using conditional transfer learning. *Frontiers in Human Neuroscience*, 11:334, 2017.

[273] Wei Lu, Haiyan Liu, Hua Ma, Tien-Ping Tan, and Lingnan Xia. Hybrid transfer learning strategy for cross-subject eeg emotion recognition. *Frontiers in Human Neuroscience*, 17, 2023.

[274] Carman Neustaedter, Saul Greenberg, and Michael Boyle. Blur filtration fails to preserve privacy for home-based video conferencing. *ACM Transactions on Computer-Human Interaction,*, pages '–36, 2006.

[275] Milena Sina Wütschert, Diana Pereira, Hartmut Schulze, and Achim Elfering. Working from home: Cognitive irritation as mediator of the link between perceived privacy and sleep problems. *Industrial Health*, 59:308–317, 10 2021.

[276] James J. Gross. Emotion regulation: Affective, cognitive, and social consequences. *Psychophysiology*, 39(3):281–291, 2002.

[277] Rodrigo Vicencio-Moreira, Regan L Mandryk, Carl Gutwin, and Scott Bateman. The effectiveness (or lack thereof) of aim-assist techniques in first-person shooter games. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 937–946, 2014.

[278] Deepak Gopinath, Siddarth Jain, and Brenna D Argall. Human-in-the-loop optimization of shared autonomy in assistive robotics. *IEEE Robotics and Automation Letters*, 2(1):247–254, 2016.

[279] Shervin Javdani, Siddhartha S Srinivasa, and J Andrew Bagnell. Shared autonomy via hindsight optimization. *Robotics Science and Systems: Online Proceedings*, 2015, 2015.

[280] Siddharth Reddy, Anca D Dragan, and Sergey Levine. Shared autonomy via deep reinforcement learning. *arXiv preprint arXiv:1802.01744*, 2018.

[281] Jonathan Liebers, Mark Abdelaziz, Lukas Mecke, Alia Saad, Jonas Auda, Uwe Gruenefeld, Florian Alt, and Stefan Schneegass. Understanding user identification in virtual reality through behavioral biometrics and the effect of body normalization. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pages 1–11, 2021.

[282] Vishal Nayak and Matthew Turk. Emotional expression in virtual agents through body language. In *International Symposium on Visual Computing*, pages 313–320. Springer, 2005.

[283] Catherine Pelachaud. Modelling multimodal expression of emotion in a virtual agent. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1535):3539–3548, 2009.

[284] Chang Yun, Dvijesh Shastri, Ioannis Pavlidis, and Zhigang Deng. O' game, can you feel my frustration?: Improving user's gaming experience via stresscam. *Conference on Human Factors in Computing Systems - Proceedings*, pages 2195–2204, 2009.