Micro-Gesture Datasets and Their Role in Gesture Recognition and Human-Computer Interaction: A Survey

www.surveyx.cn

Abstract

Micro-gesture datasets are crucial to advancing gesture recognition and human-computer interaction (HCI) technologies, offering essential data for interpreting subtle human movements and enhancing interaction intuitiveness. This survey paper emphasizes the pivotal role of these datasets in addressing user fatigue in extended reality interactions and improving static gesture classification. Challenges such as the scarcity of diverse datasets and the misclassification of gestures due to subtlety are highlighted. The survey underscores the importance of participatory design in gesture creation, advocating for user involvement to ensure intuitive system development. Technological advancements, including ultrawideband radar and electromyography sensors, are explored for their role in enhancing recognition accuracy. The integration of multi-modal and sensor fusion methods demonstrates significant improvements in understanding user intent. Innovative frameworks incorporating deep learning and attention mechanisms further advance gesture recognition, improving feature extraction and classification accuracy. The paper calls for more diverse and inclusive datasets to capture the complexity of micro-gestures, ensuring robust model generalization. Future research should focus on these challenges and explore new applications to unlock the full potential of micro-gesture recognition in creating intuitive and immersive HCI systems.

1 Introduction

1.1 Importance of Micro-Gesture Datasets

Micro-gesture datasets are essential for advancing gesture recognition and human-computer interaction (HCI), as they provide critical data for understanding subtle human actions and emotions. These datasets are particularly valuable for alleviating user fatigue in extended reality contexts by facilitating smaller, less strenuous gestures, thus enhancing user experience in augmented and virtual reality applications. Their capacity to improve the recognition of both spontaneous and nuanced movements is vital for developing intuitive interaction systems, especially in myoelectric control systems where minimizing false activations is crucial. By leveraging advanced computer vision techniques, micro-gesture datasets enhance gesture recognition, addressing challenges such as user fatigue associated with traditional input methods [1, 2, 3].

Despite their importance, the field faces challenges due to a lack of publicly available datasets focused on micro-gestures, which are crucial for analyzing complex human behaviors and understanding suppressed emotions [4]. These datasets are particularly significant for classifying static gestures, which are inherently difficult to recognize due to their lack of movement [5]. Furthermore, micro-gesture datasets improve the usability of intuitive interactions by providing structured data that supports the development of advanced HCI systems [6].

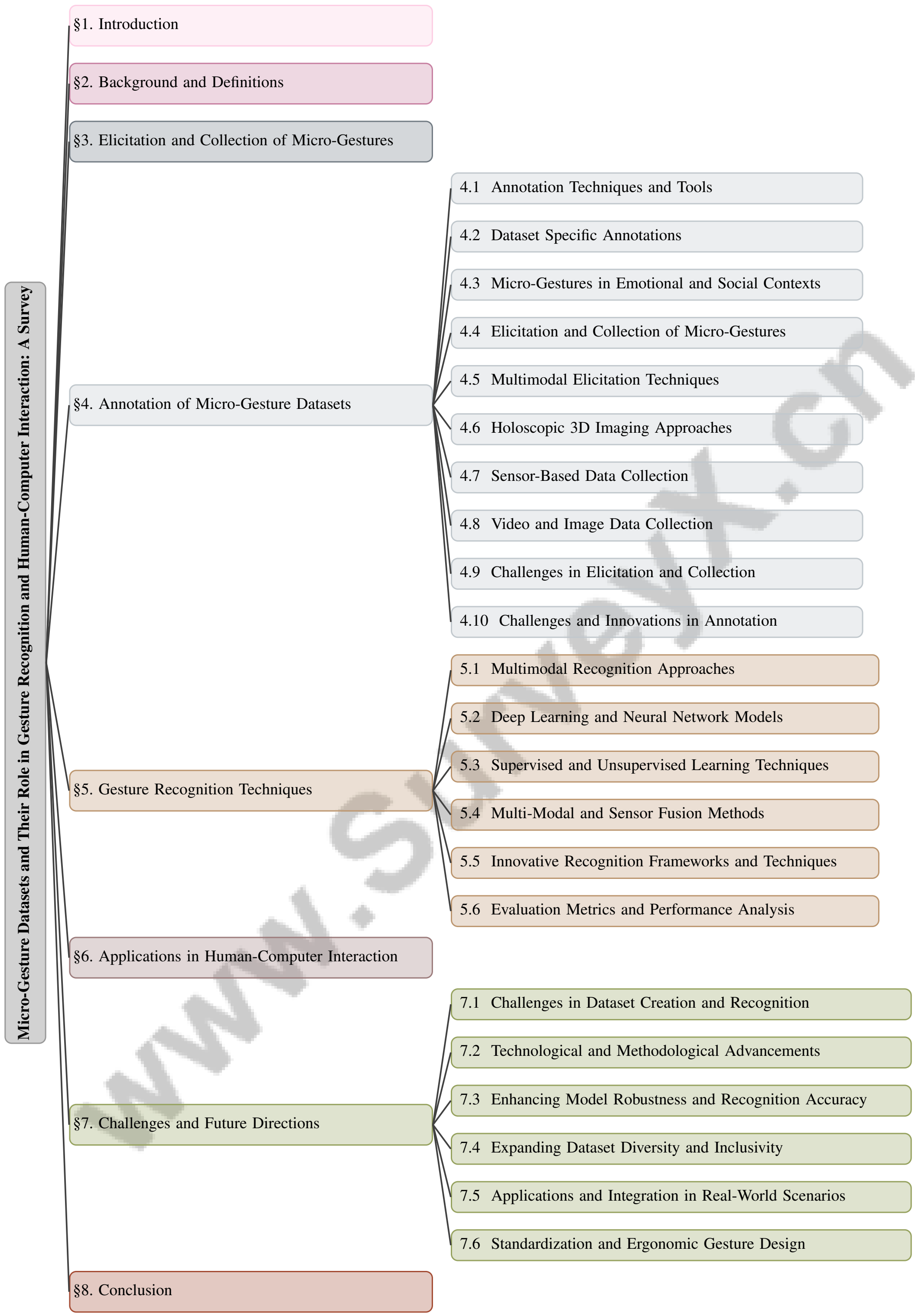

Figure 1: chapter structure

The critical role of micro-gesture datasets extends to recognizing human actions and emotions, which is fundamental for creating more natural and effective communication systems within HCI [3]. As interest in micro-gesture recognition grows across various fields, the development and refinement of these datasets remain vital for overcoming existing technological limitations and advancing gesture recognition capabilities.

1.2 Motivation for the Survey

This survey is motivated by the need to address several critical challenges and gaps in micro-gesture recognition and its application in HCI. A primary focus is on the complexities associated with mid-air hand gestures in data visualization, a field that remains underexplored despite its potential to transform interaction paradigms [6]. Additionally, the survey aims to mitigate inadvertent false activations in myoelectric control systems, which can significantly impair user experience during activities of daily living (ADLs) [7].

Another key motivation is to overcome legacy bias in gesture elicitation studies, which restricts users from proposing optimal gestures, thereby limiting the evolution of more natural gesture-based systems [8]. The survey also addresses the scarcity of comprehensive datasets in micro-gesture recognition, a deficiency that hinders advancements in augmented and virtual reality applications [4].

Furthermore, the survey seeks to resolve the missing interface problem, which arises when physical objects are enhanced with digital content, posing challenges for effective user interaction with augmented objects [9]. By thoroughly examining these issues, the survey aims to propel research forward, ensuring that micro-gesture recognition technologies continue to evolve and enhance HCI systems.

1.3 Scope of the Paper

This survey encompasses a broad range of topics critical to the advancement of micro-gesture recognition and its applications in HCI. It includes the development of benchmarks for automatic classification and online recognition of micro-gestures, with a focus on emotional recognition, which is vital for enhancing user experience [10]. The survey also explores gesture and speech interactions within augmented reality head-mounted display (AR-HMD) environments, examining timing, effectiveness, and user preferences to provide insights into multimodal interaction dynamics [11].

Additionally, the survey investigates stylus-based interactions and gesture elicitation for note-taking actions, emphasizing the mental models users develop while intentionally excluding keyboard-centric interactions to maintain a focused exploration of gesture-based methods [12]. The inclusion of a hierarchical sensor fusion approach, which combines wearable pressure sensors with Doppler radar, highlights the survey's commitment to exploring innovative methodologies that enhance gesture recognition accuracy [13].

Particularly attentive to interaction methods for augmented reality systems, the survey emphasizes micro-gestures on physical objects while intentionally excluding broader AR topics that do not involve direct interaction with physical objects or the technical details of AR content creation [9]. Moreover, it focuses on understanding user interactions with emerging technologies, advocating for robust frameworks that improve user interaction design while excluding other qualitative research methods that do not utilize object elicitation. By clearly defining these boundaries, the survey aims to provide a comprehensive and focused analysis of micro-gesture recognition and its implications for future HCI developments.

1.4 Structure of the Survey

The survey is systematically organized into several key sections, each addressing distinct facets of micro-gesture datasets and their implications for gesture recognition and HCI. The introductory section establishes the foundational importance of micro-gesture datasets, discussing their role in enhancing gesture recognition technologies and the motivation behind this comprehensive survey. Following this, the background and definitions section thoroughly defines micro-gestures and explores their significance in HCI, along with the challenges in their recognition and analysis.

The third section examines methods for eliciting and collecting micro-gesture data, highlighting multimodal elicitation techniques, holoscopic 3D imaging approaches, sensor-based data collection, and the use of video and image data, followed by a discussion on challenges encountered in these processes. The fourth section focuses on the annotation of micro-gesture datasets, emphasizing the importance of accurate annotation, common practices, tools used, and innovations in this area.

The fifth section provides an in-depth review of various gesture recognition techniques applied to micro-gesture datasets, including multimodal recognition approaches, deep learning models, and supervised and unsupervised learning techniques. It also discusses multi-modal and sensor fusion methods, innovative recognition frameworks, and evaluation metrics for performance analysis.

Applications in HCI are explored in the sixth section, where the utilization of micro-gesture recognition in software applications, advancements in gesture recognition models, real-time processing, emotion analysis, and intuitive interaction are discussed, highlighting significant achievements in gesture classification.

The penultimate section identifies current challenges in micro-gesture dataset creation and recognition, exploring potential technological and methodological advancements, strategies for enhancing model robustness and accuracy, and the need for diverse and inclusive datasets. It also discusses the potential for real-world applications and the necessity for standardization and ergonomic gesture design.

In conclusion, this paper synthesizes essential points discussed, emphasizing the pivotal role of micro-gesture datasets in enhancing gesture recognition and HCI. It highlights significant advancements achieved through the development of these datasets, including novel video datasets for intentional micro-gestures and the application of machine learning models, which illustrate the feasibility and challenges of recognizing subtle movements. Furthermore, the conclusion underscores the implications of ongoing and future research in this domain, particularly in addressing the limitations of traditional gesture recognition methods and paving the way for more intuitive and fatigue-free interaction techniques in extended reality environments [1, 14, 2, 15].The following sections are organized as shown in Figure 1.

2 Background and Definitions

2.1 Definitions and Key Concepts

Micro-gestures, comprising subtle hand and finger movements, are pivotal for interaction in augmented reality (AR) and virtual reality (VR) applications [4]. These involuntary gestures convey emotional information and reveal hidden emotions [16]. In immersive environments, they offer a natural interface, overcoming limitations of traditional input methods [6]. Recognizing micro-gestures is challenging due to their subtlety, demanding advanced detection techniques. Static gestures, or immobile hand configurations, are particularly difficult to recognize due to orientation and anatomical variations [5]. Benchmarks for finger micro-gestures like Button, Dial, and Slider are critical for interactive control [17]. Micro-gestures also emerge from user-generated input proposals via participatory design, differing from expert-defined gestures [18]. Defining these concepts aids in creating robust systems for recognizing and utilizing micro-gestures, enhancing human-computer interaction and intuitive communication systems.

2.2 Significance in Human-Computer Interaction

Micro-gestures enhance human-computer interaction (HCI) by enabling natural interaction methods, especially in AR and VR environments [18]. Their subtlety allows seamless integration into daily activities, essential for precise control and minimal fatigue [7]. In VR, they personalize interactions, improving accessibility for users with diverse capabilities, including motor impairments [19]. Micro-gestures address challenges in interacting with physical objects augmented with digital content, crucial for effective mixed-reality user interfaces [9]. They expand interaction possibilities in human-robot interaction (HRI), especially for non-humanoid robots, enhancing robots' ability to interpret human actions [20]. Integrating micro-gestures with speech in multimodal HRI datasets captures explicit and implicit insights, advancing the field [21]. In affective computing, micro-gestures enrich emotion recognition datasets, offering a rich source of gestural behaviors that enhance understanding of emotional states [22, 23]. Datasets like iMiGUE provide benchmarks for emotion recognition while ensuring privacy and inclusivity [24]. The significance of micro-gestures in HCI is underscored by the need for controlled experiments to minimize biases in AI systems, ensuring equitable interactions across diverse user groups [25]. Comprehensive frameworks for evaluating human-object interaction models foster advancements in AR and VR, reinforcing the critical role of micro-gestures in HCI evolution [26].

2.3 Challenges in Recognition and Analysis

The recognition and analysis of micro-gestures face significant challenges that hinder effective gesture recognition systems. A major obstacle is the imbalanced distribution of gesture categories, complicating classification and leading to intra-class variability with minimal inter-class differences [16]. This variability can cause misclassification of ambiguous samples, affecting recognition accuracy. Existing benchmarks are limited by data availability and accuracy, particularly when using traditional imaging systems like RGB-D cameras, which may not capture the nuances required for micro-gesture recognition [17]. Lack of standardization in mid-air hand gestures further complicates recognition, limiting effectiveness across visualization contexts [6]. This issue is exacerbated by inefficiencies in analyzing gesture proposals during elicitation studies, where manual classification predominates [27]. As users generate more symbols, gesture proposal diversity decreases, hampering the identification of novel gestures and innovative interaction techniques [8]. Existing benchmarks suffer from poor gesture capture quality and lack comprehensive datasets, issues that databases like HoMG aim to address by providing higher fidelity data [4]. Researchers face challenges related to detection fidelity and spatial resolution in current systems, which do not support subtle micro-gestures, reducing their effectiveness in practical applications [9]. Addressing these challenges requires developing advanced recognition techniques and comprehensive datasets to accurately capture micro-gestures. Progress in micro-gesture recognition is crucial for enhancing HCI, especially as extended reality technologies become more prevalent. These gestures can reduce fatigue associated with traditional gestures, promoting comfortable interactions over prolonged use. Developing datasets for intentional micro-gestures and applying machine learning models to detect subtle movements will facilitate more natural user interactions, particularly in XR headsets, enhancing user experience and enabling innovative gesture-based interfaces [2, 12].

2.4 Micro-Gestures in Emotional and Social Contexts

Micro-gestures are crucial in emotional and social interactions, serving as subtle yet powerful conveyors of non-verbal communication. These involuntary movements provide insights into emotional states, enhancing understanding of social dynamics and relationships. Studies highlight their significance in conveying hidden emotions more reliably than facial expressions, necessitating a multidisciplinary approach to analyze these gestures [1, 10, 28, 3]. In affective computing, micro-gestures are integral to systems recognizing and interpreting emotions through non-verbal cues, allowing detection of concealed emotions and offering comprehensive user interaction understanding in virtual and physical environments.

As illustrated in Figure 4, which depicts the role of micro-gestures in different contexts, these gestures are essential not only for emotional recognition but also for social interactions. The figure underscores the challenges faced in the development of algorithms that can accurately interpret these subtle cues. In social contexts, micro-gestures enrich emotional expression, acting as indicators of agreement, disagreement, or interest, complementing verbal communication and enhancing interaction experiences. Their subtlety allows them to function in social exchanges, influencing conversation flow and social cue perception [22].

Integrating micro-gestures into HCI systems can create empathetic and responsive technologies. By leveraging emotional and social dimensions, HCI systems can intuitively respond to user needs, fostering engaging and personalized interactions [21]. This is particularly relevant in VR and AR environments, where interpreting emotional cues enhances user immersion and satisfaction. The significance of micro-gestures in emotional and social contexts underscores the need for robust datasets and recognition systems capable of capturing these subtle movements. Advancements in emotion recognition technologies will enhance precision, enabling machines to interpret emotions through non-verbal cues, including micro-gestures and body language. Improved accuracy will facilitate the development of intuitive communication systems leveraging multimodal data, enriching human interaction in digital and physical environments, fostering deeper connections and meaningful exchanges [24, 22, 3, 12, 28].

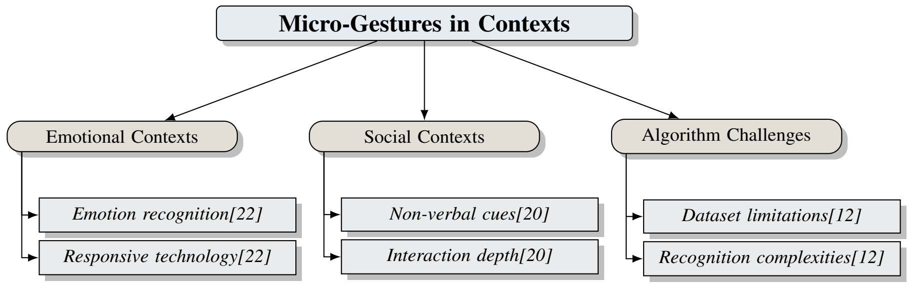

Figure 2: This figure illustrates the role of micro-gestures in different contexts, highlighting their importance in emotional recognition, social interactions, and the challenges faced in algorithm development.

3 Elicitation and Collection of Micro-Gestures

Understanding human-computer interaction (HCI) requires effective elicitation and collection of micro-gestures, crucial for developing gesture recognition systems. This section explores methodologies for capturing these subtle movements, emphasizing multimodal elicitation techniques as a primary approach. By integrating diverse data sources, researchers can enrich datasets, fostering intuitive and responsive HCI technologies. The subsequent subsection details specific multimodal elicitation techniques that have significantly advanced micro-gesture recognition.

3.1 Multimodal Elicitation Techniques

Multimodal elicitation techniques are essential in HCI research for capturing the intricate nature of micro-gestures. By combining modalities such as video, motion capture, and sensor data, these techniques provide a comprehensive representation of micro-gestures crucial for advanced gesture recognition systems. Participatory design methods, exemplified by the Unlimited Production Gesture Elicitation (UPGE) approach, encourage participants to generate numerous gestures for each referent, facilitating the exploration and refinement of gesture sets [8].

The integration of multiple modalities enhances dataset richness and applicability. GestureMap, an advanced visual analytics tool, utilizes a Variational Autoencoder to project 3D skeletal representations onto an interactive 2D map, allowing researchers to explore gesture patterns and user preferences. This tool employs techniques like DTW Barycenter Averaging for quick visual representation and variance assessment, aiding in clustering with k-means and deriving consensus sets of user-defined gestures [27, 12].

Sensor technologies are crucial in multimodal elicitation. Techniques such as UWB-based Static Gesture Recognition (UWB-SGR) leverage UWB radar technology to capture detailed data on static gestures, showcasing radar's potential for fine-grained gesture information. The On-Demand Myoelectric Control (ODMC) method combines myoelectric signals with gesture inputs, enhancing recognition accuracy by minimizing false activations during daily activities. This method allows users to switch control modes using wake gestures, filtering out unrelated muscle movements while maintaining sensitivity for intentional inputs [11, 7, 29, 9].

The use of Microsoft Kinect cameras exemplifies the integration of diverse modalities, capturing authentic micro-gestures and generating synthetic ones via advanced 3D modeling software. This approach enriches the dataset for training recognition models and enhances the accuracy of gesture recognition systems, particularly in immersive applications like virtual and augmented reality (VR and AR). Holoscopic 3D imaging technology provides multi-viewpoint images that offer depth and spatial information, addressing the limitations of traditional 2D imaging methods and significantly improving micro-gesture recognition performance in HCI [1, 2, 15, 30].

As illustrated in Figure 5, the hierarchical categorization of multimodal elicitation techniques in HCI research emphasizes the importance of participatory design methods, sensor technologies, and temporal modeling approaches in enhancing micro-gesture recognition systems. Temporal modeling techniques, such as the Bidirectional Long Short-Term Memory Variational Autoencoder (bLSTM-

VAE), effectively capture dynamics in sequential skeleton data, highlighting the importance of temporal modeling in multimodal elicitation.

Advancements in multimodal elicitation techniques enhance recognition systems' accuracy and robustness, supporting more intuitive HCI technologies. As micro-gesture recognition progresses, research increasingly focuses on robust methods for identifying and interpreting subtle movements, particularly in extended HCI contexts within immersive environments. This focus addresses challenges such as user fatigue with traditional gestures and the need for comprehensive datasets for training machine learning models. Recent studies have introduced novel datasets and methodologies to improve micro-gesture detection accuracy, underscoring the critical role of this research in enhancing user experience in gesture-based interfaces [14, 2, 15].

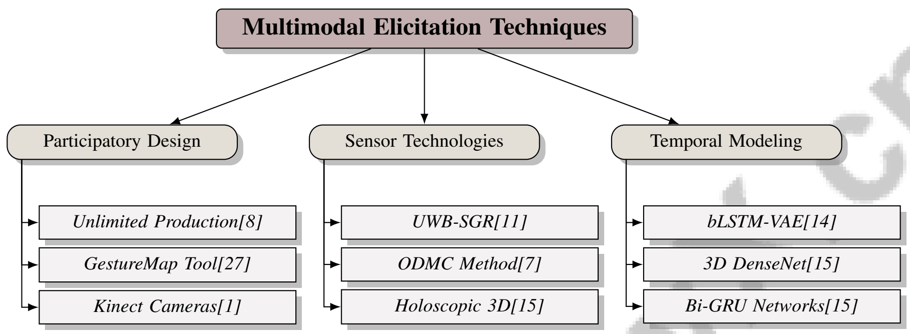

Figure 3: This figure illustrates the hierarchical categorization of multimodal elicitation techniques in HCI research, focusing on participatory design methods, sensor technologies, and temporal modeling approaches to enhance micro-gesture recognition systems.

3.2 Holoscopic 3D Imaging Approaches

Holoscopic 3D (H3D) imaging has become a significant technique for capturing micro-gestures, enhancing gesture recognition system precision and accuracy. This technology uses integral imaging principles to capture three-dimensional information, akin to human stereoscopic vision, providing a comprehensive representation of subtle human gestures [15].

H3D imaging systems excel in capturing finger micro-gestures with high accuracy, vital for gesture recognition and HCI applications [17]. The ability of H3D imaging to capture depth information alongside spatial and temporal data enables nuanced analysis of micro-gestures, often characterized by subtle movements challenging to discern with traditional 2D imaging techniques.

The application of holoscopic 3D imaging in micro-gesture dataset collection enhances the precision of finger micro-gesture capture, a crucial advantage for robust gesture recognition system development [15]. Integrating advanced imaging technologies is pivotal in overcoming challenges associated with recognizing subtle and complex micro-gestures, characterized by minimal movement and intricate details [17].

As research in micro-gesture recognition evolves, adopting holoscopic 3D imaging technologies presents a promising avenue for significantly improving the fidelity and accuracy of micro-gesture datasets. Holoscopic 3D cameras, utilizing a microlens array to capture depth and multi-angle information in a single 2D image, overcome traditional 2D sensor limitations. Recent advancements, including comprehensive holoscopic micro-gesture databases and advanced deep learning algorithms, have demonstrated substantial improvements in recognition accuracy, making these technologies vital for enhancing HCI in immersive applications such as virtual and augmented reality [4, 31, 15, 30]. The development of these advanced imaging techniques is essential for advancing HCI and unlocking new possibilities for intuitive and immersive interaction technologies.

3.3 Sensor-Based Data Collection

Sensor-based data collection is crucial for acquiring micro-gesture datasets, providing the precision and detail necessary for effective gesture recognition systems. Various sensors, including inertial measurement units (IMUs), electromyography (EMG) sensors, and depth sensors, capture the subtle nuances of micro-gestures, often characterized by minimal movement and intricate detail. IMUs are particularly valuable, tracking motion dynamics with high accuracy, essential for capturing micro-gesture details [4].

EMG sensors enhance micro-gesture data collection by measuring muscle activity, offering insights into the physiological aspects of gesture execution. This is particularly beneficial in myoelectric control applications, where understanding underlying muscle movements can improve gesture recognition accuracy and responsiveness [7]. Integrating EMG data with other sensor modalities facilitates a comprehensive understanding of micro-gestures, promoting the development of robust and intuitive HCI technologies [16].

Depth sensors, such as those in Microsoft Kinect, provide additional capabilities by capturing spatial information crucial for distinguishing subtle variations in gesture execution [4]. Combining depth data with motion and muscle activity information offers a multi-faceted view of micro-gestures, enabling more accurate and reliable recognition systems [17].

Implementing sensor-based data collection techniques in micro-gesture recognition significantly enhances dataset accuracy and richness, facilitating sophisticated recognition algorithms capable of managing the inherent complexity and variability of micro-gestures. This advancement is critical for HCI applications, where precise identification and timing of micro-gestures can improve user experience and communication [1, 14, 2]. As the field evolves, integrating diverse sensor technologies will remain a key research area, driving advancements in gesture recognition system accuracy and applicability in HCI.

3.4 Video and Image Data Collection

Collecting micro-gestures through video and image data is essential for creating comprehensive datasets for gesture recognition systems. Video data captures the dynamic and temporal aspects of micro-gestures, revealing the subtlety and fluidity of movements crucial for accurate recognition [15]. High-resolution video recordings enable detailed observations of gesture execution, allowing for the analysis of intricate hand and finger movements often imperceptible through other modalities.

Image data provides static snapshots for analyzing specific postures and configurations of micro-gestures. High-quality images facilitate the examination of fine details, such as finger positioning and hand orientation, critical for distinguishing similar gestures [4]. Integrating image data with video recordings enriches the dataset, offering both temporal and spatial information invaluable for training robust gesture recognition models.

Advanced imaging techniques, such as those in holoscopic 3D imaging, augment the collection process by capturing depth information alongside traditional video and image data [15]. This additional data layer allows for a more comprehensive understanding of the spatial dynamics involved in micro-gesture execution, particularly beneficial for augmented and virtual reality applications where depth perception is critical [17].

Integrating video and image data in micro-gesture datasets significantly enhances gesture recognition system accuracy and reliability, enabling the development of more intuitive and immersive HCI technologies. This advancement addresses traditional gesture limitations, which can lead to user fatigue, by facilitating the recognition of smaller, less physically demanding micro-gestures, promoting longer and more comfortable interactions with computer systems in extended reality environments [1, 2]. As research progresses, refining video and image data collection methods will be pivotal in advancing gesture-based system capabilities.

3.5 Challenges in Elicitation and Collection

Eliciting and collecting micro-gesture data presents several challenges that hinder the development of comprehensive gesture recognition systems. A primary obstacle is the diversity of user experiences and contextual factors influencing gesture usage, necessitating the design of functional gestures representative of users' capabilities and preferences. This diversity highlights the importance of a varied participant pool to capture a wide range of interaction preferences and mitigate legacy bias, which can limit innovative gesture-based system evolution [18].

Another significant challenge is the reliance on high-quality training data and accurately annotated datasets, as the precision of gesture recognition systems heavily depends on these factors [1]. The tedious nature of manual gesture analysis complicates the elicitation process, limiting scalability and repeatability of studies and hindering effective exploration of the gesture space [27]. This is compounded by traditional methods' limitations in property elicitation, emphasizing the need for innovative approaches to overcome these obstacles [32].

Ensuring accurate performance across segmented interaction spaces and addressing user confidence in gesture execution are additional challenges that must be resolved to enhance gesture recognition systems' reliability and effectiveness [33]. Existing datasets often focus on elementary tasks, revealing limitations in scaling to more intricate domains and prioritizing human command data over robot behavior records, which hampers the development of comprehensive datasets suitable for advanced applications [21].

To effectively address these challenges, developing innovative elicitation methodologies and refining data collection techniques is essential. This approach will ensure the creation of robust, diverse, and accurately annotated micro-gesture datasets, crucial for enhancing gesture recognition systems' performance across various applications, including intelligent note-taking interfaces and emotion recognition from micro-gestures, thereby facilitating more intuitive HCI and improving user experience [1, 12, 8].

4 Annotation of Micro-Gesture Datasets

The annotation process is a cornerstone of developing micro-gesture datasets, ensuring their precision and reliability for gesture recognition systems. This section delves into the techniques and tools integral to annotating micro-gestures, highlighting the complexities of accurately labeling these subtle movements, and sets the stage for discussing specific annotation methodologies.

4.1 Annotation Techniques and Tools

Annotation is pivotal for creating effective gesture recognition models by accurately labeling subtle gestures, which serve as foundational data for machine learning algorithms. Techniques like equivalence classes, which categorize similar gestures, streamline annotation and enhance model accuracy [16, 8]. Recent advances, such as machine learning algorithms for automated annotation, have improved dataset precision and reduced manual labeling efforts [3]. However, manual annotation remains subjective, prompting the need for standardized protocols to ensure consistency [8, 6]. Tools like GestureMap integrate visualization with computational analysis, aiding in gesture data clustering [27]. Privacy concerns, particularly regarding personal data, are also critical during data collection [25]. The integration of video and sensor-based data necessitates sophisticated algorithms for synchronization and analysis [17]. Overall, developing innovative annotation techniques, especially those using machine learning, is crucial for enhancing the efficiency and accuracy of dataset annotation [3].

4.2 Dataset Specific Annotations

Each micro-gesture dataset requires tailored annotations to serve its specific application. For instance, the iMiGUE dataset focuses on emotion recognition and requires annotations capturing emotional expressions through subtle movements [24]. In contrast, datasets for static gestures like Button, Dial, and Slider need annotations considering precise hand configurations [17]. Variability in gesture data due to individual and environmental factors complicates annotation [16]. Researchers are leveraging machine learning models to automatically detect and classify micro-gestures, enhancing dataset robustness [23]. The rise of AR and VR environments further necessitates comprehensive annotation guidelines to ensure consistency across datasets [18].

4.3 Micro-Gestures in Emotional and Social Contexts

Micro-gestures play a crucial role in conveying emotions and intentions, enhancing communication in HCI by facilitating nuanced interaction. In social contexts, these gestures serve as non-verbal cues, adding depth to interactions [20]. In affective computing, recognizing micro-gestures is key to improving emotion recognition systems, enabling responsive technology [22]. Micro-gestures offer a more efficient alternative to traditional gestures, reducing user fatigue. However, the lack of comprehensive datasets for intentional micro-gestures hinders algorithm development. Recent studies highlight the feasibility of detecting these subtle movements and the complexities of refining their recognition, particularly in extended reality environments [14, 2, 12].

As illustrated in Figure 4, the role of micro-gestures spans various contexts, emphasizing their significance in emotional recognition and social interactions, while also shedding light on the challenges faced in algorithm development. Developing robust algorithms to distinguish micro-gestures remains a significant research focus, essential for enhancing HCI.

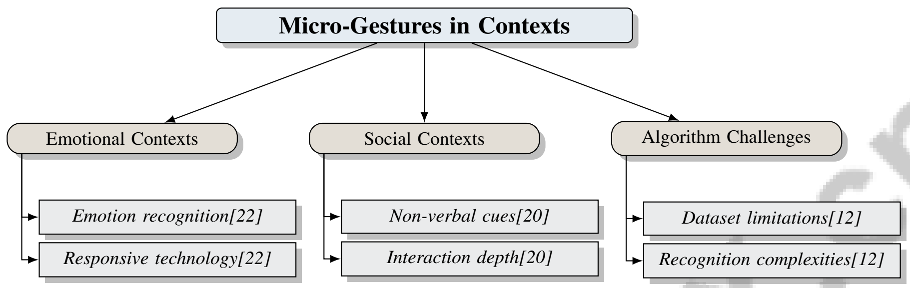

Figure 4: This figure illustrates the role of micro-gestures in different contexts, highlighting their importance in emotional recognition, social interactions, and the challenges faced in algorithm development.

4.4 Elicitation and Collection of Micro-Gestures

Eliciting and collecting micro-gesture data are foundational for creating extensive datasets, improving gesture recognition systems' performance, particularly in understanding human emotions and HCI applications [1, 14]. Various methods have been explored to capture micro-gestures, each with unique challenges.

4.5 Multimodal Elicitation Techniques

Multimodal elicitation techniques are prominent for collecting micro-gesture data, using multiple input modalities to capture gesture complexity. These techniques often involve sensors like cameras and accelerometers to create comprehensive datasets, enhancing recognition accuracy by addressing single-modality limitations [13]. Recent interest in multimodal approaches highlights their potential to improve system robustness by leveraging complementary sensor information [21]. Combining depth sensors with wearable sensors enhances subtle finger movement detection [13]. However, traditional methods often rely on predefined gestures that may not align with users' natural styles, leading to engagement issues [8].

Participatory design approaches involving users in creating gesture vocabularies are increasingly emphasized [18]. As illustrated in Figure 5, the hierarchical categorization of multimodal elicitation techniques in HCI research focuses on participatory design methods, sensor technologies, and temporal modeling approaches to enhance micro-gesture recognition systems. High precision in capturing subtle movements often requires advanced, costly hardware [4]. Variability in gesture execution necessitates adaptive algorithms, complicating data collection [16]. Overcoming these challenges is crucial for advancing micro-gesture recognition systems [12, 11, 8].

4.6 Holoscopic 3D Imaging Approaches

Holoscopic 3D imaging is pivotal in collecting micro-gesture data, capturing intricate details of subtle movements. This technology mimics human stereoscopic vision to capture three-dimensional information, providing comprehensive gesture representation [15]. The capability to capture depth, spatial, and temporal data is advantageous for recognizing micro-gestures [17]. Integrating holoscopic 3D imaging improves gesture recognition precision, especially in AR and VR applications [22]. However, specialized equipment and expertise are needed, potentially limiting accessibility [4]. Processing high-dimensional data requires advanced computational techniques, complicating data

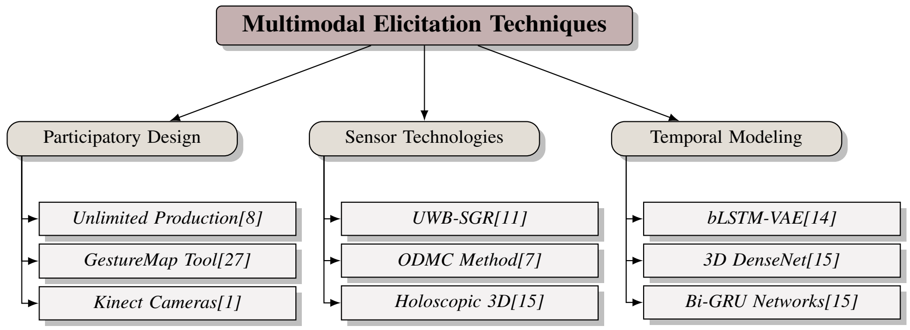

Figure 5: This figure illustrates the hierarchical categorization of multimodal elicitation techniques in HCI research, focusing on participatory design methods, sensor technologies, and temporal modeling approaches to enhance micro-gesture recognition systems.

collection [16]. Current research focuses on developing cost-effective solutions and robust methods for processing intricate data [31, 15, 30]. Addressing these challenges is essential for advancing micro-gesture recognition in HCI.

4.7 Sensor-Based Data Collection

Sensor-based data collection is fundamental for micro-gesture recognition, providing the precision necessary for robust systems. Various sensors, including IMUs, EMG sensors, and depth sensors, capture subtle nuances of micro-gestures [4]. IMUs track motion dynamics with high accuracy, essential for capturing micro-gesture details [4]. EMG sensors measure muscle activity, offering insights into gesture execution's physiological aspects, crucial for myoelectric control applications [7]. Integrating EMG data with other sensors fosters comprehensive understanding, facilitating robust HCI technologies [16]. Depth sensors capture spatial information critical for distinguishing subtle gesture variations [4]. Combining depth, motion, and muscle activity data offers a multifaceted view, enabling accurate recognition systems [17]. Sensor-based data collection enhances dataset accuracy, supporting sophisticated recognition algorithms, crucial for HCI applications [1, 14, 2]. Continued research into sensor technologies will drive advancements in recognition accuracy and applicability.

4.8 Video and Image Data Collection

Video and image data collection is critical for creating comprehensive micro-gesture datasets. Video captures dynamic and temporal aspects, revealing subtle movements essential for accurate recognition [15]. High-resolution recordings enable detailed gesture observation, facilitating analysis of intricate hand and finger movements. Image data provides static snapshots for analyzing specific postures, enabling examination of fine details crucial for distinguishing similar gestures [4]. Integrating image data with video enriches the dataset, providing temporal and spatial information invaluable for training robust models. Advanced imaging techniques, like holoscopic 3D imaging, augment collection by capturing depth alongside traditional data [15]. This enhances understanding of spatial dynamics, beneficial for AR and VR applications [17]. Combining video and image data improves recognition system accuracy, supporting intuitive HCI technologies. Enhancing video and image data collection techniques is crucial for improving gesture-based systems' accuracy and reliability, essential for recognizing human actions and emotions in real-time [2, 3].

4.9 Challenges in Elicitation and Collection

Eliciting and collecting micro-gesture data presents challenges that hinder comprehensive gesture recognition system development. A primary obstacle is user diversity and contextual factors influencing gesture usage, necessitating functional gesture designs reflective of users' capabilities and preferences. This diversity underscores the need for varied participant pools to capture interaction preferences and mitigate legacy bias [18]. Another challenge is the reliance on high-quality training data and accurately annotated datasets, as system precision heavily depends on these factors [1]. Manual gesture analysis's tedious nature complicates elicitation, limiting scalability and repeatability, hindering gesture space exploration [27]. Traditional methods' limitations in property elicitation emphasize the need for innovative approaches [32]. Ensuring accurate performance across segmented spaces and addressing user confidence in gesture execution are additional challenges [33]. Existing datasets often focus on elementary tasks, revealing limitations in scaling to intricate domains and prioritizing human command data over robot behavior records, hindering comprehensive dataset development for advanced applications [21]. Developing innovative elicitation methodologies and refining data collection techniques are essential to address these challenges, ensuring robust, diverse, and accurately annotated micro-gesture datasets, crucial for enhancing gesture recognition systems [1, 12, 8].

4.10 Challenges and Innovations in Annotation

Annotating micro-gesture datasets is complex, presenting challenges critical for effective gesture recognition systems. Subjectivity in manual annotation can lead to inconsistencies, affecting recognition models' generalization capabilities [32]. Accurate and consistent annotations are needed to address gesture diversity and contextual influences. Innovations in annotation techniques, like automated machine learning algorithms, enhance accuracy and efficiency by reducing manual labeling time and increasing consistency [1, 14, 2]. Visualization tools, such as GestureMap, provide interactive interfaces for visualizing and analyzing gesture data, enabling pattern exploration and consensus derivation [32]. Integrating multiple data modalities, like video, image, and sensor inputs, remains a challenge. Sophisticated algorithms for synchronizing and analyzing diverse data types are essential for comprehensive annotations. The subtlety of micro-gestures necessitates high-fidelity data and precise annotations, crucial for distinguishing nuanced movements, enhancing recognition systems' effectiveness in applications from HCI to emotional analysis [3, 12, 1, 2, 14]. Ongoing advancements in annotation techniques are vital for addressing gesture recognition complexities. Recent studies emphasize eliciting diverse and effective gestures for applications like intelligent note-taking and micro-gesture recognition. Refining these techniques enhances detection accuracy and improves user interactions, leading to more intuitive gesture-based interfaces [14, 3, 12, 8]. Ensuring annotation accuracy and consistency will enhance recognition model training and evaluation, contributing to more effective HCI technologies.

5 Gesture Recognition Techniques

5.1 Multimodal Recognition Approaches

Multimodal recognition approaches are crucial for enhancing the accuracy and robustness of gesture recognition systems by integrating diverse input sources such as hand gestures, speech, and sensor data. These techniques foster intuitive user experiences in augmented reality (AR) environments by combining gesture and speech interactions, which is essential for immersive applications [7]. Recent advancements in computer vision models, including VideoMAE, MViT, 3D ResNet, and C3D, have significantly improved micro-gesture recognition accuracy by capturing intricate human movements [2]. When incorporated into multimodal systems, these models create a robust framework for recognizing complex gestures in dynamic settings. The bLSTM-VAE method exemplifies the potential of multimodal approaches by learning robust representations from noisy data, enhancing action recognition [3].

Innovative systems like the On-Demand Myoelectric Control (ODMC) demonstrate the efficacy of multimodal approaches in real-time applications, using wake gestures to improve accuracy and minimize false activations [7]. This integration accommodates diverse user needs, particularly for individuals with specific physical capabilities.

Research in multimodal recognition techniques is vital for advancing human-computer interaction. By leveraging multiple modalities, researchers can better analyze and respond to user intent and actions. Studies reveal that gestures often precede speech by 81 milliseconds, indicating a close alignment of information conveyed through both modalities. Identifying user-defined gestures through extensive elicitation studies refines gesture sets, enhancing recognition efficiency and facilitating natural interactions [11, 29, 8].

5.2 Deep Learning and Neural Network Models

Deep learning and neural network models have significantly advanced micro-gesture recognition by improving feature extraction and classification accuracy. The integration of 3D-CNN-based recognition networks, using skeletal and semantic embedding losses, has notably enhanced gesture recognition accuracy [16]. Convolutional neural networks (CNNs) with attention mechanisms refine feature extraction by focusing on salient features, addressing the challenges posed by the subtlety and variability of micro-gestures.

The synergy between advanced sensor technologies and deep learning frameworks, such as UWB radar technology integrated with CNN and MobileNet, markedly enhances static gesture recognition accuracy [5]. The iMAP framework further advances natural gesture recognition, exemplifying innovative deep learning approaches [34]. Models trained on diverse demographic data demonstrate deep learning's adaptability across varied contexts, utilizing vector splits for comprehensive evaluation [25].

The continuous evolution of deep learning models underscores their critical role in advancing human-computer interaction technologies. By employing advanced algorithms and integrating multimodal data, these models enhance interaction system design, fostering more natural and efficient user experiences [29, 35, 12, 11, 18].

5.3 Supervised and Unsupervised Learning Techniques

Both supervised and unsupervised learning techniques play crucial roles in gesture recognition, enabling systems to interpret micro-gestures effectively. Supervised learning techniques, such as support vector machines (SVMs) and decision trees, require comprehensive and accurately annotated datasets to classify micro-gestures by learning from labeled examples [16, 23]. Conversely, unsupervised learning techniques, including clustering algorithms and self-organizing maps, identify patterns within data without explicit labels, offering valuable insights in scenarios with scarce labeled data or new gesture sets [8].

A hybrid approach combining supervised learning for initial training with unsupervised learning for continuous adaptation enhances system robustness and adaptability, improving generalization across diverse users and environments [16]. Semi-supervised learning, which uses both labeled and unlabeled data, further enhances generalization across diverse scenarios, enabling robust micro-gesture recognition [3]. Investigating these learning techniques is crucial for advancing gesture recognition technology, allowing machines to accurately interpret human actions and emotions through robust methods such as skeleton-based recognition and multi-scale graph convolution networks [1, 3].

5.4 Multi-Modal and Sensor Fusion Methods

Multi-modal and sensor fusion methods are pivotal in enhancing micro-gesture recognition systems by integrating data from various sensors, such as cameras and inertial measurement units (IMUs). These methods provide a nuanced representation of micro-gestures, which are subtle movements crucial for expressing emotions and facilitating communication [1, 14, 2]. By integrating EMG signals with video data, researchers achieve a more comprehensive understanding of the physiological and kinematic aspects of micro-gestures, improving recognition accuracy [14, 2, 3].

Sensor fusion approaches, such as combining depth sensors and motion capture systems, enhance gesture recognition performance. UWB-based Static Gesture Recognition (UWB-SGR) technology, for example, demonstrates significant potential in recognizing subtle gestures [17]. By leveraging multiple sensor technologies, researchers develop robust recognition models capable of handling the complexity and variability of micro-gestures [16].

Integrating multimodal data not only improves gesture recognition accuracy but also supports the development of more intuitive human-computer interaction (HCI) technologies. As the field evolves, refining sensor-based data collection methods remains critical. Addressing micro-gesture recognition complexities contributes to developing more intuitive interaction systems, enhancing user experience across various applications, particularly in extended reality (XR) environments [30, 12, 9, 2, 14].

5.5 Innovative Recognition Frameworks and Techniques

Advancements in gesture recognition have led to innovative frameworks and techniques that enhance the accuracy and efficiency of recognizing micro-gestures. These methods leverage cutting-edge machine learning algorithms and advanced sensor technologies to address challenges posed by the subtlety and variability of micro-gestures. Studies highlight the limitations of traditional gestures, such as fatigue from mid-air movements, proposing micro-gestures as a more efficient alternative [12, 9, 1, 2, 14].

Innovative frameworks incorporate deep learning models with attention mechanisms to improve feature extraction and classification processes, enhancing recognition accuracy [17]. Hybrid models that combine supervised and unsupervised learning techniques enhance gesture recognition system adaptability and robustness, leveraging both paradigms to improve generalization across diverse users and environments [16, 3].

Sensor fusion methodologies have been instrumental in advancing gesture recognition frameworks. By integrating data from diverse modalities, including video recordings and other sensor technologies, researchers achieve a nuanced understanding of micro-gestures, which are subtle, often unconscious movements that convey emotional states. This multimodal approach enhances recognition accuracy and robustness, facilitating better differentiation between various gestures and improving performance in emotional analysis and human-computer interaction tasks [24, 1, 36, 2, 14].

Ongoing investigation and enhancement of innovative recognition frameworks are essential for advancing human-computer interaction technologies. By addressing challenges such as noise resistance in skeleton data extraction and accurate gesture detection in dynamic environments, researchers contribute to developing more intuitive and effective interaction systems that enhance user experience across various applications [14, 2, 3].

5.6 Evaluation Metrics and Performance Analysis

Benchmark	Size	Domain	Task Format	Metric
HoMG[4]	31,595	Gesture Recognition	Micro-Gesture Recognition	Accuracy, F1-score
iMiGUE[24]	18,499	Emotion Analysis	Micro-Gesture Recognition	Accuracy
DESK[37]	1,286	Surgical Robotics	Surgeme Classification	Accuracy
SMG[10]	821,056	Psychology	Gesture Classification	Accuracy, F1-score
MiGA[14]	3,692	Micro-gesture Recognition	Action Recognition	F1 score
MiGA[23]	22,191	Behavior Analysis	Micro-gesture Classification	Top-1 accuracy, F1
iReplica[26]	680,000	Human-Object Interaction	Contact Detection	Average Precision, Accuracy
MMAD[36]	6,528	Micro-Action	Multi-label Action Detection	Detection-mAP

Table 1: This table presents a comprehensive overview of various benchmarks utilized in micro-gesture recognition and related domains. It details the size, domain, task format, and evaluation metrics for each dataset, highlighting the diversity and scope of current research efforts in human-computer interaction. The benchmarks include datasets from gesture recognition, emotion analysis, surgical robotics, and human-object interaction, providing a broad spectrum for evaluating model performance.

Evaluating micro-gesture recognition systems is crucial for developing robust technologies that enhance user experience by reducing physical strain compared to traditional gestures, particularly in extended reality environments [14, 2]. Metrics such as accuracy, precision, recall, and F1-score provide insights into model recognition capabilities. Accuracy, representing the proportion of correctly classified instances over the total number of instances, remains a fundamental metric. Advanced imaging techniques, such as holoscopic 3D imaging, have demonstrated potential in improving micro-gesture recognition precision, achieving notable accuracy improvements [4].

Precision and recall are critical for evaluating model performance in scenarios with class imbalance. Precision measures the proportion of true positive instances among all instances classified as positive, while recall assesses the proportion of true positive instances among all actual positive instances. These metrics are essential for understanding trade-offs between false positives and false negatives in recognition systems.

Recent advancements in gesture recognition techniques have demonstrated significant improvements in performance metrics. For example, a proposed method achieved accuracies of 47.9% on the

Spontaneous Micro-Gesture (SMG) dataset and 37.5% on the Micro-Gesture (MG) dataset [4]. Cross-subject evaluation protocols, which involve dividing subjects into training and testing groups, are critical for performance analysis, ensuring models are evaluated on their ability to generalize across different users, enhancing their applicability in real-world scenarios [24]. Table 1 provides a detailed overview of representative benchmarks in micro-gesture recognition and related fields, illustrating the diversity of datasets and evaluation metrics employed in recent research.

In recent years, the field of human-computer interaction (HCI) has witnessed significant advancements, particularly in the area of micro-gesture recognition. The hierarchical structure of applications in this domain is crucial for understanding the various components that contribute to the efficacy of gesture recognition systems. As illustrated in Figure 6, this figure categorizes the impact on software applications and highlights key technological developments, including advancements in gesture recognition models, real-time processing capabilities, emotion analysis, and achievements in gesture classification. By delineating these aspects, the figure underscores the contributions of each element to the overall enhancement of HCI systems, thereby providing a comprehensive overview of the current landscape in micro-gesture recognition.

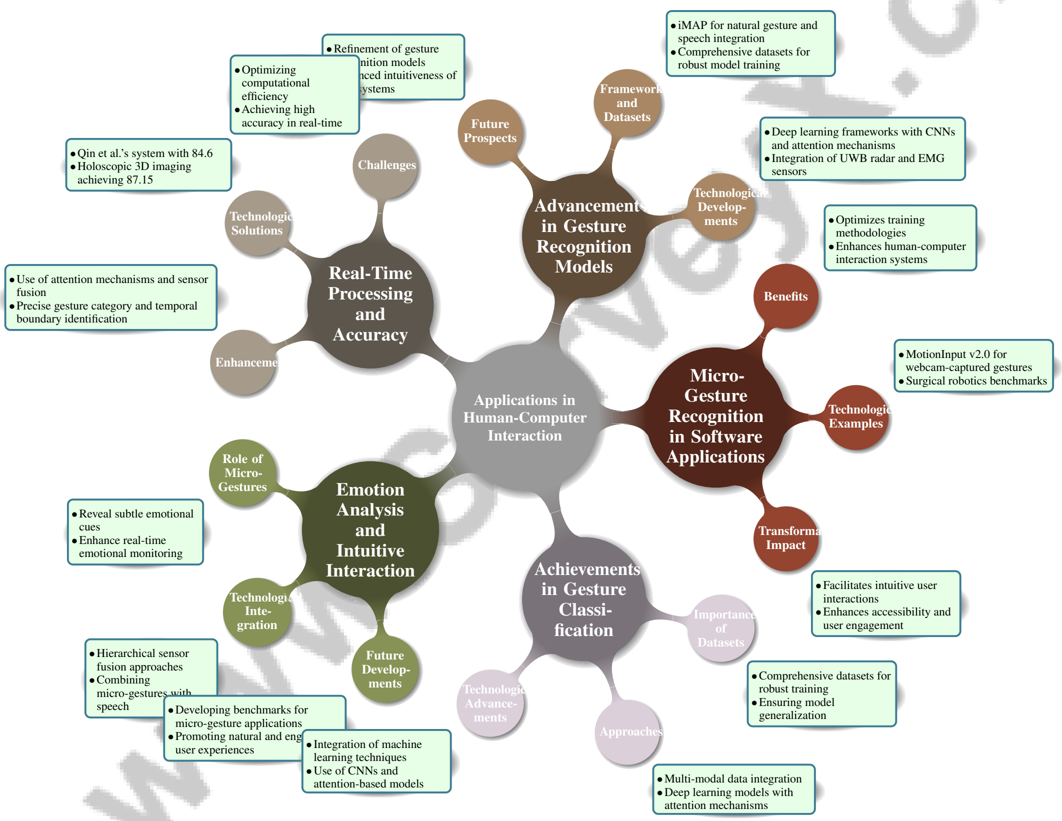

Figure 6: This figure illustrates the hierarchical structure of applications in human-computer interaction, focusing on micro-gesture recognition. It categorizes the impact on software applications, advancements in gesture recognition models, real-time processing, emotion analysis, and achievements in gesture classification, highlighting key technological developments and their contributions to enhancing HCI systems.

6 Applications in Human-Computer Interaction

6.1 Micro-Gesture Recognition in Software Applications

Micro-gesture recognition has significantly transformed software applications by facilitating intuitive user interactions. MotionInput v2.0 exemplifies this by utilizing webcam-captured gestures for seamless application control, enhancing accessibility and user engagement beyond traditional input devices [38]. In surgical robotics, benchmarks aid in knowledge transfer across platforms, optimizing training methodologies and enhancing human-computer interaction systems [37, 2]. The high accuracy and real-time processing capabilities of these systems position them as pivotal in shaping future HCI technologies [5].

6.2 Advancements in Gesture Recognition Models

Recent developments in gesture recognition models have notably enhanced the accuracy and efficiency of micro-gesture recognition, crucial for intuitive HCI. Deep learning frameworks, particularly those incorporating convolutional neural networks (CNNs) and attention mechanisms, have improved feature extraction and classification by focusing on key features within gesture data [17]. The integration of advanced sensor technologies, such as ultrawideband (UWB) radar and electromyography (EMG) sensors, with deep learning models has significantly advanced static gesture recognition accuracy [5, 7]. Frameworks like iMAP facilitate seamless interactions by integrating natural gestures with speech [34]. Comprehensive datasets capturing diverse micro-gestures ensure robust model training and evaluation, promoting generalization across varied user demographics [24]. As the field evolves, these technological integrations will continue to refine gesture recognition models, enhancing HCI systems' intuitiveness.

6.3 Real-Time Processing and Accuracy

Real-time processing and accuracy are critical for effective micro-gesture recognition systems, especially in HCI applications requiring immediate feedback. Achieving these involves optimizing computational efficiency while accurately identifying subtle micro-gestures, a challenge highlighted by our success in the Micro-gesture Online Recognition track at the MiGA challenge during IJCAI 2024 [14, 2, 12, 9]. Although deep learning models' computational demands can impede real-time deployment, recent advancements like Qin et al.'s system have achieved 84.6

6.4 Emotion Analysis and Intuitive Interaction

Micro-gestures are pivotal in emotion analysis, revealing subtle emotional cues and enhancing intuitive interactions. Their subtlety allows for the recognition of emotional states often missed by traditional methods, thereby improving real-time emotional monitoring [1]. Integrating micro-gestures into emotion analysis frameworks has significantly advanced emotional state recognition and interpretation, fostering empathetic HCI systems [3]. Hierarchical sensor fusion approaches enhance gesture recognition accuracy, enabling the detection of subtle emotional cues [13]. Furthermore, combining micro-gestures with modalities like speech enhances model performance, aligning interactions with natural communication patterns, particularly in AR and VR environments [21, 17]. Developing benchmarks for micro-gesture applications is critical for improving recognition accuracy and robustness, ultimately contributing to more intuitive HCI systems [23, 16]. As research progresses, integrating micro-gestures into emotion analysis will significantly enhance HCI technologies, promoting natural and engaging user experiences.

6.5 Achievements in Gesture Classification

Advancements in gesture classification have substantially improved recognition systems' accuracy and efficiency, particularly in HCI contexts. The integration of machine learning techniques, notably CNNs and attention-based models, has enhanced micro-gesture classification accuracy by capturing intricate gesture features [17]. Sophisticated frameworks utilizing sensor fusion techniques, which combine data from sources like UWB radar and pressure sensors, have shown to increase recognition accuracy by approximately 15

7 Challenges and Future Directions

7.1 Challenges in Dataset Creation and Recognition

The development of micro-gesture recognition systems is hindered by several challenges in dataset creation and recognition. A major issue is the limited availability and diversity of datasets, which restricts the development of robust models. The lack of publicly available micro-gesture datasets and the imbalanced distribution of gesture categories within existing datasets lead to substantial intra-class variability, complicating effective classification [4, 16]. High-quality skeleton data is crucial for accurate pose estimation; inaccuracies can result in misclassification, particularly in noisy environments [3, 9]. Variations in angles and distances further complicate static gesture recognition [5]. Additionally, the small size of video subsets in current datasets limits the development of robust video-based models [17]. Legacy biases in gesture design limit innovation, and overlapping feature spaces between intentional gestures and typical muscle activations can result in false activations [6, 7]. The reliance on specific gestures and limited participant diversity within datasets affects generalizability, necessitating more inclusive data collection methodologies [4]. Tools like GestureMap can help address these challenges by providing comprehensive visualization and analysis tools for large datasets [27]. Future research should focus on innovative data collection methodologies and refined recognition techniques to improve dataset robustness and diversity, paving the way for more intuitive human-computer interactions [2, 3, 12].

7.2 Technological and Methodological Advancements

Advancements in micro-gesture recognition hinge on technological innovations and methodological refinements. Integrating additional modalities, such as eye-gazing, can improve interaction efficiency and alleviate fatigue associated with mid-air gestures [29]. Addressing cultural differences in gesture design is crucial for developing universally accepted systems [12]. Enhanced machine learning methodologies, such as deep learning with multi-observation loss functions and augmentation strategies, can improve predictions in complex applications like micro-gesture recognition [30, 3, 32, 25]. Exploring the impact of varying object types on interaction proposals, particularly in multimodal scenarios, can yield insights into multimodal communication dynamics [11]. Refinements in participant engagement during elicitation studies are crucial for advancing micro-gesture recognition methodologies [35]. Embedded sensors in wearable devices, such as earables, offer a promising avenue for improving gesture recognition in real-world scenarios [33]. Incorporating physics-based models and real-time environment reconstruction into frameworks could enhance benchmark applicability [26]. Future research should focus on mitigating legacy bias and expanding elicitation methodologies, improving clustering accuracy, integrating live data analysis, and expanding tools like GestureMap to support various contexts [27]. Enhancing detection accuracy through feature selection and validation techniques will further advance the field [9].

7.3 Enhancing Model Robustness and Recognition Accuracy

Enhancing gesture recognition model robustness and accuracy involves optimizing the balance between false positives and negatives, as seen in systems like On-Demand Myoelectric Control (ODMC), which can improve reliability [7]. Data augmentation techniques can enhance model robustness by simulating diverse scenarios, aligning with the need for effective emotion recognition systems [35, 22, 32, 18]. Platforms like Magic Xroom advance our understanding of user interactions and emotional responses. Exploring new applications, such as the iMAP framework, demonstrates gesture recognition's potential in creating intuitive interaction systems [34]. Developing sophisticated testing environments is crucial for evaluating and refining models, ensuring effective function in dynamic environments [17]. Continuous exploration and refinement of methodologies, coupled with advanced sensor technologies and data augmentation, will advance model robustness and accuracy.

7.4 Expanding Dataset Diversity and Inclusivity

Expanding dataset diversity and inclusivity is vital for advancing micro-gesture recognition systems, impacting generalizability across diverse populations. Diverse datasets capture a wide range of micro-gestural behaviors, ensuring robust models for real-world scenarios [18]. Current datasets often lack demographic diversity, leading to biases and reduced applicability [21]. Efforts to collect data from diverse demographics, including variations in age, gender, ethnicity, and abilities, are necessary [18]. Capturing contextual factors influencing micro-gesture execution, such as environmental conditions and cultural differences, is increasingly recognized [21]. Inclusive datasets address accessibility challenges, ensuring systems accommodate diverse physical and cognitive abilities [19].

7.5 Applications and Integration in Real-World Scenarios

Micro-gesture recognition systems have the potential to enhance HCI across diverse domains, enabling intuitive communication through subtle gestures. In AR and VR, these systems enhance user experience by providing natural interaction methods [14, 2, 30]. In healthcare, they offer promising applications in assistive devices for individuals with motor impairments, improving accessibility [7, 19]. In smart environments and IoT, micro-gesture systems facilitate seamless interaction with connected devices, enhancing convenience and efficiency [17]. In collaborative work environments, they enhance virtual meeting effectiveness by enabling natural communication [21, 11]. In automotive technology, they enhance driver safety and comfort by enabling hands-free control [5].

7.6 Standardization and Ergonomic Gesture Design

Developing micro-gesture recognition systems requires a focus on standardization and ergonomic gesture design. Standardization is crucial for creating consistent gesture vocabularies that can be universally understood across platforms [18]. The absence of standardized gesture sets can lead to fragmentation and user confusion [8]. Researchers are exploring approaches to develop universal vocabularies and comprehensive guidelines for gesture design [18]. Ergonomic gesture design is critical, ensuring gestures are comfortable and do not contribute to fatigue [7]. This is essential in applications involving extended interactions, such as VR and AR [19]. Emphasizing ergonomic aspects ensures HCI systems are effective, user-friendly, and accessible [18]. The need for standardization is underscored by the growing diversity of HCI applications, requiring consistent interaction methods across platforms [25].

8 Conclusion

Micro-gesture datasets play a pivotal role in enhancing gesture recognition and human-computer interaction (HCI) technologies. By enabling systems to accurately interpret subtle human movements, these datasets contribute to more intuitive and effective interaction technologies. The HoMG database, for instance, sets a benchmark for future research, showcasing the potential of micro-gesture recognition to address challenges such as user fatigue in extended reality interactions and the misclassification of static gestures.

The survey underscores the significance of involving users in the gesture design process to ensure systems are intuitive and aligned with user preferences. Methods like the Unlimited Production Gesture Elicitation (UPGE) enable the exploration and refinement of gesture sets, fostering more natural interaction techniques. Moreover, the integration of advanced sensor technologies, such as ultrawideband (UWB) radar and electromyography (EMG) sensors, has markedly enhanced the accuracy and robustness of micro-gesture recognition systems. Multi-modal and sensor fusion methods, which combine data from various sensors, have substantially improved recognition performance, offering a deeper understanding of user intent and actions.

Advancements in recognition frameworks, particularly those utilizing deep learning models with attention mechanisms, have significantly propelled gesture recognition forward. These models enhance feature extraction and classification accuracy, enabling precise recognition of micro-gestures and advancing HCI technologies. The necessity for comprehensive and accurately annotated datasets that capture the diversity and complexity of human interactions is highlighted by benchmarks like the Multi-label Micro-Action Detection (MMAD) benchmark, which encourages further research into co-occurring micro-actions and their implications for gesture recognition systems.

Future research should focus on overcoming challenges in dataset creation and recognition, emphasizing the development of diverse and inclusive datasets that mirror the variability and complexity of micro-gestures. Exploring new applications and integration strategies for micro-gesture recognition systems will be crucial for advancing the field and unlocking new opportunities for intuitive and immersive interaction technologies. By prioritizing the development of diverse datasets, researchers can enhance the generalization capabilities of recognition models, ensuring their applicability across various user populations and interaction contexts. The findings underscore the significant potential of micro-gestures as indicators of hidden emotional states, highlighting the importance of continued exploration in this domain.

References

[1] Haoyu Chen. Human gesture and micro-gesture analysis: datasets, methods, and applications. 2022.

[2] Nikhil Krishnaswamy and Nathaniel Blanchard. Intentional microgesture recognition for extended human-computer interaction.

[3] Henglin Shi. Learning-based human action and affective gesture analysis. 2023.

[4] Yi Liu, Hongying Meng, Mohammad Rafiq Swash, Yona Falinie A Gaus, and Rui Qin. Holoscopic 3d micro-gesture database for wearable device interaction. In *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*, pages 802–807. IEEE, 2018.

[5] Abhishek Sebastian and Pragna R. A technique for classifying static gestures using uwb radar, 2024.

[6] Zona Kostic, Catherine Dumas, Sarah Pratt, and Johanna Beyer. Exploring mid-air hand interaction in data visualization, 2023.

[7] Ethan Eddy, Evan Campbell, Scott Bateman, and Erik Scheme. On-demand myoelectric control using wake gestures to eliminate false activations during activities of daily living, 2024.

[8] Andreea Danielescu and David Piorkowski. Iterative design of gestures during elicitation: Understanding the role of increased production, 2022.

[9] Klen Čopič Pucihar, Christian Sandor, Matjaž Kljun, Wolfgang Huerst, Alexander Plopski, Takafumi Taketomi, Hirokazu Kato, and Luis A Leiva. The missing interface: Micro-gestures on augmented objects. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–6, 2019.

[10] Haoyu Chen, Henglin Shi, Xin Liu, Xiaobai Li, and Guoying Zhao. Smg: A micro-gesture dataset towards spontaneous body gestures for emotional stress state analysis. *International Journal of Computer Vision*, 131(6):1346–1366, 2023.

[11] Adam S. Williams and Francisco R. Ortega. Understanding gesture and speech multimodal interactions for manipulation tasks in augmented reality using unconstrained elicitation, 2022.

[12] Katy Ilonka Gero, Lydia B. Chilton, Chris Melancon, and Mike Cleron. Eliciting gestures for novel note-taking interactions, 2021.

[13] Haobo Li, Xiangpeng Liang, Aman Shrestha, Yuchi Liu, Hadi Heidari, Julien Le Kernec, and Francesco Fioranelli. Hierarchical sensor fusion for micro-gesture recognition with pressure sensor array and radar. *IEEE Journal of Electromagnetics, RF and Microwaves in Medicine and Biology*, 4(3):225–232, 2019.

[14] Pengyu Liu, Fei Wang, Kun Li, Guoliang Chen, Yanyan Wei, Shengeng Tang, Zhiliang Wu, and Dan Guo. Micro-gesture online recognition using learnable query points. *arXiv preprint arXiv:2407.04490*, 2024.

[15] Yi Liu, Shuang Yang, Hongying Meng, Mohammad Rafiq Swash, and Shiguang Shan. A novel pseudo viewpoint based holoscopic 3d micro-gesture recognition. In *Companion Publication of the 2020 International Conference on Multimodal Interaction*, pages 77–81, 2020.

[16] Kun Li, Dan Guo, Guoliang Chen, Xinge Peng, and Meng Wang. Joint skeletal and semantic embedding loss for micro-gesture classification. *arXiv preprint arXiv:2307.10624*, 2023.

[17] Min Peng, Chongyang Wang, and Tong Chen. Attention based residual network for micro-gesture recognition. In *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*, pages 790–794. IEEE, 2018.

[18] Adam S. Williams and Francisco R. Ortega. A concise guide to elicitation methodology, 2021.

[19] Jingze Tian, Yingna Wang, Keye Yu, Liyi Xu, Junan Xie, Franklin Mingzhe Li, Yafeng Niu, and Mingming Fan. Designing upper-body gesture interaction with and for people with spinal muscular atrophy in vr, 2024.

[20] Jan Leusmann, Steeven Villa, Thomas Liang, Chao Wang, Albrecht Schmidt, and Sven Mayer. An approach to elicit human-understandable robot expressions to support human-robot interaction, 2024.

[21] Snehesh Shrestha, Yantian Zha, Saketh Banagiri, Ge Gao, Yiannis Aloimonos, and Cornelia Fermuller. Natsgd: A dataset with speech, gestures, and demonstrations for robot learning in natural human-robot interaction. *arXiv preprint arXiv:2403.02274*, 2024.

[22] S. M. Hossein Mousavi, Matteo Besenzoni, Davide Andreoletti, Achille Peternier, and Silvia Giordano. The magic xroom: A flexible vr platform for controlled emotion elicitation and recognition, 2024.

[23] Haoyu Chen, Björn W Schuller, Ehsan Adeli, and Guoying Zhao. The 2nd challenge on micro-gesture analysis for hidden emotion understanding (miga) 2024: Dataset and results. In *MiGA 2024: Proceedings of IJCAI 2024 Workshop&Challenge on Micro-gesture Analysis for Hidden Emotion Understanding (MiGA 2024) co-located with 33rd International Joint Conference on Artificial Intelligence (IJCAI 2024)*. RWTH Aachen, 2024.

[24] Xin Liu, Henglin Shi, Haoyu Chen, Zitong Yu, Xiaobai Li, and Guoying Zhao. imigue: An identity-free video dataset for micro-gesture understanding and emotion analysis. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10631–10642, 2021.

[25] Keeley Crockett, James O'Shea, and Wasiq Khan. Automated deception detection of males and females from non-verbal facial micro-gestures. In *2020 International Joint Conference on Neural Networks (IJCNN)*, pages 1–7. IEEE, 2020.

[26] Vladimir Guzov, Julian Chibane, Riccardo Marin, Yannan He, Yunus Saracoglu, Torsten Sattler, and Gerard Pons-Moll. Interaction replica: Tracking human-object interaction and scene changes from human motion, 2024.

[27] Hai Dang and Daniel Buschek. Gesturemap: Supporting visual analytics and quantitative analysis of motion elicitation data by learning 2d embeddings, 2021.

[28] Haoyu Chen, Xin Liu, Xiaobai Li, Henglin Shi, and Guoying Zhao. Analyze spontaneous gestures for emotional stress state recognition: A micro-gesture dataset and analysis with deep learning. In *2019 14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019)*, pages 1–8. IEEE, 2019.

[29] Xiaoyan Zhou, Adam S. Williams, and Francisco R. Ortega. Eliciting multimodal gesture+speech interactions in a multi-object augmented reality environment, 2022.

[30] Yi Liu, Min Peng, Mohammad Rafiq Swash, Tong Chen, Rui Qin, and Hongying Meng. Holoscopic 3d microgesture recognition by deep neural network model based on viewpoint images and decision fusion. *IEEE Transactions on Human-Machine Systems*, 51(2):162–171, 2021.

[31] Rui Qin, Yi Liu, Mohammad Rafiq Swash, Maozhen Li, Hongying Meng, Tao Lei, and Tong Chen. A fast automatic holoscopic 3d micro-gesture recognition system for immersive applications. In *Advances in Natural Computation, Fuzzy Systems and Knowledge Discovery: Volume 2*, pages 696–703. Springer, 2020.

[32] Sebastian Casalaina-Martin, Rafael Frongillo, Tom Morgan, and Bo Waggoner. Multi-observation elicitation, 2017.

[33] Shaikh Shawon Arefin Shimon, Ali Neshati, Junwei Sun, Qiang Xu, and Jian Zhao. Exploring uni-manual around ear off-device gestures for earables, 2024.

[34] S. Kettebekov and R. Sharma. Toward natural gesture/speech control of a large display, 2001.

[35] Carla Willig. Reflections on the use of object elicitation. *Qualitative Psychology*, 4(3):211, 2017.

[36] Kun Li, Dan Guo, Pengyu Liu, Guoliang Chen, and Meng Wang. Mmad: Multi-label micro-action detection in videos. *arXiv preprint arXiv:2407.05311*, 2024.

[37] Naveen Madapana, Md Masudur Rahman, Natalia Sanchez-Tamayo, Mythra V. Balakuntala, Glebys Gonzalez, Jyothsna Padmakumar Bindu, L. N. Vishnunandan Venkatesh, Xingguang Zhang, Juan Barragan Noguera, Thomas Low, Richard Voyles, Yexiang Xue, and Juan Wachs. Desk: A robotic activity dataset for dexterous surgical skills transfer to medical robots, 2019.

[38] Ashild Kummen, Guanlin Li, Ali Hassan, Teodora Ganeva, Qianying Lu, Robert Shaw, Chenuka Ratwatte, Yang Zou, Lu Han, Emil Almazov, Sheena Visram, Andrew Taylor, Neil J Sebire, Lee Stott, Yvonne Rogers, Graham Roberts, and Dean Mohamedally. Motioninput v2.0 supporting directx: A modular library of open-source gesture-based machine learning and computer vision methods for interacting and controlling existing software with a webcam, 2021.

Disclaimer:

SurveyX is an AI-powered system designed to automate the generation of surveys. While it aims to produce high-quality, coherent, and comprehensive surveys with accurate citations, the final output is derived from the AI's synthesis of pre-processed materials, which may contain limitations or inaccuracies. As such, the generated content should not be used for academic publication or formal submissions and must be independently reviewed and verified. The developers of SurveyX do not assume responsibility for any errors or consequences arising from the use of the generated surveys.