DOCTORAL THESIS

Motion-based Feature Analysis for the Design of Full-body Interactions in the Context of Computer Vision & Large Volume Spaces

Author:
Antonio ESCAMILLA

Supervisors:
Dr. Javier MELENCHÓN
Dr. Carlos MONZO

A thesis submitted in fulfillment of the requirements for the Doctor of Philosophy

in the

Doctoral Programme in Networks and Information Technologies
IT, Multimedia and Telecommunications Department
Doctoral School

October 27, 2023

Acknowledgements

I would like to express my sincere gratitude and appreciation to my supervisors, Dr. Javier Melenchón and Dr. Carlos Monzo, for their support, guidance, and expertise throughout my doctoral journey. Their valuable insights and encouragement have played a key role in shaping this thesis. I am truly fortunate to have had the opportunity to work under their mentorship. I am also immensely grateful to Dr. Jose A. Morán and Dr. Juan Pablo Carrascal for their contributions to my research over the past four years. Their collaborative efforts, constructive feedback, and intellectual discussions have greatly enriched the quality and depth of my work. Their dedication and willingness to share their knowledge have been instrumental in my growth as a researcher.

I would like to express my deep gratitude to the Universitat Oberta de Catalunya for the doctoral grant that supported my research endeavors. Additionally, I am grateful to the university for welcoming me as a pre-doctoral researcher in the vibrant city of Barcelona. The nurturing environment, resources, and academic community provided by the university have been instrumental in shaping my academic and personal growth.

Furthermore, I extend my appreciation to the interaction design professionals from various countries who generously participated in the focus groups and interviews. Their willingness to share their expertise and experiences has been invaluable in broadening my understanding of their daily work strategies. I am sincerely thankful for their time, insights, and contributions, which have significantly enriched the empirical foundation of this thesis.

Finally, my heartfelt thanks go to my family, my beloved wife *Maria Paula*, and our beautiful children *Martin and Amaia*, for their love, understanding, and support throughout this demanding journey. Their patience, encouragement, and belief in me have been my constant source of motivation. I would also like to extend my gratitude to my parents for their endless love, sacrifices, and faith in my abilities. Their support has been the foundation upon which I have built my academic pursuits.

Contents

Abstract

Human-computer interaction (HCI) is evolving towards more natural, social, and user-centered interactions, specifically focusing on the use of the entire body as an interface. Full-body interaction involves capturing and analyzing body movements to control and manipulate digital content, enabling immersive and engaging experiences. In this context, computer vision technology is preferred over add-ons, wearables, or marker systems due to its non-intrusive, cost-effective, and versatile nature. By eliminating the need for additional devices, it allows for a more natural and unrestricted human motion observation, ensuring a genuine representation of movement. However, designing effective full-body interaction systems requires understanding human motion and its interpretation, which poses challenges for interaction designers.

This thesis addresses the critical need for structured relationships between human motion and feature extraction technology, which arises from the existing problem that non-technically skilled practitioners often struggle to utilize motion-based features effectively. To provide better context, it is essential to present the challenges faced by these professionals in working with complex technical parameters and motion-based features, hindering their ability to create multi-user interactive experiences. By highlighting this issue early on, the thesis aims to emphasize its significance and underscore that its primary objective is not just to resolve the interaction design challenges but also to offer an additional value of empowering non-technical profiles by facilitating their design work through accessible motion feature extraction technology.

In line with the identified challenges, this thesis aims to contribute to the field of interaction design by offering a comprehensive framework for motion-based interaction design in large-volume spaces. These spaces, characterized by their size and complexity, present unique challenges in terms of seamless and intuitive user experiences. By bridging the gap between technical parameters and interpretable motion-based features, the presented framework is suited to a broader range of practitioners. The main stages that form this contribution are outlined as follows:

1. A comprehensive investigation of interaction designers' perceptions of motion-based full-body features and their link to sensor-based interaction design strategies. This study provides insight into designers' perspectives, experiences, and challenges when incorporating motion-based features into their design work.

2. Development of a guiding framework to align sensor-based interactions with the design process, capturing the essential requirements to use motion-based features and empowering designers to create more engaging and meaningful interactions.

3. A set of motion-based features carefully developed in alignment with the established framework. This stage underscores the significance of the framework as a guiding principle in shaping the design of user interactions in large-volume spaces. This symbiotic relationship between the proposed features and the framework highlights a valuable contribution, enabling non-technologically skilled practitioners to create immersive and engaging multi-user interactive experiences.

4. A user-centered evaluation to assess the impact of the proposed features on concept ideation and motion-based interaction design. This evaluation gathers insights from users and designers, providing valuable feedback on the effectiveness and usability of the proposed features.

Resumen

La interacción humano-computadora (HCI) está evolucionando hacia interacciones más naturales, sociales y centradas en el usuario, centrándose específicamente en el uso de todo el cuerpo como interfaz. La interacción de cuerpo completo implica capturar y analizar los movimientos del cuerpo para controlar y manipular contenido digital, lo que permite experiencias inmersivas y atractivas. En este contexto, la tecnología de visión por computadora es preferida sobre complementos, dispositivos portátiles o sistemas de marcadores debido a su naturaleza no intrusiva, rentable y versátil. Al eliminar la necesidad de dispositivos adicionales, permite una observación más natural y sin restricciones del movimiento humano, asegurando una representación genuina del movimiento. Sin embargo, el diseño de sistemas de interacción de cuerpo completo efectivos requiere comprender el movimiento humano y su interpretación, lo que plantea desafíos para los diseñadores de interacción.

Esta tesis aborda la necesidad crítica de relaciones estructuradas entre el movimiento humano y la tecnología de extracción de características, que surge del problema existente de que los profesionales no técnicos a menudo tienen dificultades para utilizar eficazmente las características basadas en el movimiento. Para proporcionar un mejor contexto, es esencial presentar los desafíos que enfrentan estos profesionales al trabajar con parámetros técnicos complejos y características basadas en el movimiento, lo que dificulta su capacidad para crear experiencias interactivas para varios usuarios. Al resaltar este problema desde el principio, la tesis tiene como objetivo enfatizar su importancia y subrayar que su objetivo principal no es solo resolver los desafíos del diseño de interacción, sino también ofrecer un valor adicional al capacitar a perfiles no técnicos mediante la facilitación de su trabajo de diseño a través de la tecnología accesible de extracción de características basadas en el movimiento.

Siguiendo los desafíos identificados, esta tesis tiene como objetivo contribuir al campo del diseño de interacción ofreciendo un marco integral para el diseño de interacción basado en el movimiento en espacios de gran volumen. Estos espacios, caracterizados por su tamaño y complejidad, presentan desafíos únicos en términos de experiencias de usuario fluidas e intuitivas. Al cerrar la brecha entre los parámetros técnicos y las características basadas en el movimiento interpretables, el marco presentado es adecuado para una amplia gama de profesionales. Las principales etapas que conforman esta contribución se describen a continuación:

1. Una investigación exhaustiva de las percepciones de los diseñadores de interacción sobre las características basadas en el movimiento de cuerpo completo y su vínculo con las estrategias de diseño de interacción basadas en sensores. Este estudio proporciona una visión de las perspectivas, experiencias y desafíos de los diseñadores al incorporar características basadas en el movimiento en su trabajo de diseño.

2. Desarrollo de un marco guía para alinear las interacciones basadas en sensores con el proceso de diseño, capturando los requisitos esenciales para utilizar características basadas en el movimiento y empoderando a los diseñadores para crear interacciones más atractivas y significativas.

3. Un conjunto de características basadas en el movimiento desarrolladas cuidadosamente en línea con el marco establecido. Esta etapa subraya la importancia

del marco como un principio rector en la configuración del diseño de interacciones de usuario en espacios de gran volumen. Esta relación simbiótica entre las características propuestas y el marco destaca una contribución valiosa, que capacita a los profesionales no tecnológicamente capacitados para crear experiencias interactivas inmersivas y atractivas para varios usuarios.

4. Una evaluación centrada en el usuario para evaluar el impacto de las características propuestas en la ideación de conceptos y el diseño de interacción basado en el movimiento. Esta evaluación recopila información de los usuarios y diseñadores, proporcionando comentarios valiosos sobre la efectividad y la usabilidad de las características propuestas.

Resumen

La interacció humà-ordinador (HCI) està evolucionant cap a interaccions més naturals, socials i centrades en l'usuari, centrant-se específicament en l'ús de tot el cos com a interfície. La interacció de cos complet implica capturar i analitzar els moviments del cos per controlar i manipular contingut digital, cosa que permet experiències immersives i atractives. En aquest context, la tecnologia de visió per ordinador és preferida sobre complements, dispositius portàtils o sistemes de marcadors a causa de la seva naturalesa no intrusiva, rendible i versàtil. En eliminar la necessitat de dispositius addicionals, permet una observació més natural i sense restriccions del moviment humà, assegurant una representació genuïna del moviment. Tot i això, el disseny de sistemes d'interacció de cos complet efectius requereix comprendre el moviment humà i la seva interpretació, cosa que planteja desafiaments per als dissenyadors d'interacció.

Aquesta tesi aborda la necessitat crítica de relacions estructurades entre el moviment humà i la tecnologia d'extracció de característiques, que sorgeix del problema existent que els professionals no tècnics sovint tenen dificultats per utilitzar eficaçment les característiques basades en el moviment. Per proporcionar un millor context, és essencial presentar els desafiaments que enfronten aquests professionals en treballar amb paràmetres tècnics complexos i característiques basades en el moviment, cosa que en dificulta la capacitat per crear experiències interactives per a diversos usuaris. En ressaltar aquest problema des del principi, la tesi té com a objectiu emfatitzar la seva importància i subratllar que el seu objectiu principal no és sols resoldre els desafiaments del disseny d'interacció, sinó també oferir un valor addicional en capacitar perfils no tècnics mitjançant la facilitació del seu treball de disseny a través de la tecnologia accessible d'extracció de característiques basades en el moviment.

Seguint els desafiaments identificats, aquesta tesi té com a objectiu contribuir al camp del disseny d'interacció oferint un marc integral per al disseny d'interacció basat en el moviment en espais de gran volum. Aquests espais, caracteritzats per la seva mida i la seva complexitat, presenten desafiaments únics en termes d'experiències d'usuari fluïdes i intuïtives. En tancar la bretxa entre els paràmetres tècnics i les característiques basades en el moviment interpretables, el marc presentat és adequat per a una àmplia gamma de professionals. Les principals etapes que conformen aquesta contribució es d'escriuen a continuació:

1. Una investigació exhaustiva de les percepcions dels dissenyadors d'interacció sobre les característiques basades en el moviment de cos complet i el seu enllaç amb les estratègies de disseny d'interacció basades en sensors. Aquest estudi proporciona una visió de les perspectives, experiències i desafiaments dels dissenyadors en incorporar característiques basades en el moviment en el treball de disseny.

2. Desenvolupament d'un marc guia per alinear les interaccions basades en sensors amb el procés de disseny, capturant els requisits essencials per utilitzar característiques basades en el moviment i empoderant els dissenyadors per crear interaccions més atractives i significatives.

3. Un conjunt de característiques basades en el moviment desenvolupades amb cura en línia amb el marc establert. Aquesta etapa subratlla la importància del marc com a principi rector en la configuració del disseny d'interaccions

d'usuari en espais de gran volum. Aquesta relació simbiòtica entre les característiques proposades i el marc destaca una contribució valuosa, que capacita els professionals no tecnològicament capacitats per crear experiències interactives immersives i atractives per a diversos usuaris.

4. Una avaluació centrada en l'usuari per avaluar l'impacte de les característiques proposades en la ideació de conceptes i el disseny d'interacció basat en el moviment. Aquesta avaluació recopila informació dels usuaris i dissenyadors i proporciona comentaris valuosos sobre l'efectivitat i la usabilitat de les característiques proposades.

List of Figures

List of Tables

List of Abbreviations

3DPS	**3D P**ictorial **S**tructures
CLI	**C**ommand **L**ine **I**nterface
COCO	**C**ommon **O**bjects in **CO**ntext
DBSCAN	**D**ensity-**B**ased **S**patial **C**lustering of **A**pplications with **N**oise
DFD	**D**iscrete **F**réchet **D**istance
DTW	**D**ynamic **T**ime **W**arping
EDR	**E**dit **D**istance for **R**eal sequences
FD	**F**réchet **D**istance
GUI	**G**raphical **U**ser **I**nterface
HCI	**H**uman **C**omputer **I**nteraction
HMM	**H**idden **M**arkov **M**odels
IML	**I**nteractive **M**achine **L**earning
LCSS	**L**ongest **C**ommon **S**ub**S**equence
ML	**M**achine **L**earning
NUI	**N**atural **U**ser **I**nterface
SDK	**S**oftware **D**evelopment **K**it
WIMP	**W**indow, **I**cons, **M**enus and **P**ointer
YOLOv3	**Y**ou **O**nly **L**ook **O**nce, Version 3

Chapter 1

Introduction

Human-computer interaction (HCI) is a field of study that deals with the design, evaluation, and implementation of interactive computing systems for human use. HCI focuses on the relationship between people and technology and seeks to understand how to create user-friendly and effective computer interfaces that meet the needs of users [1]. Current trends in the field are oriented toward more natural, social, and user-centered interactions, leaving behind the central interest in the physical machine and the traditional confines of the desk [2]–[4]. One emergent approach in the field of HCI is the use of the entire body as an interface, which has gained momentum in recent years as researchers explore new ways to incorporate gesture recognition and other forms of non-verbal communication into computing interfaces [5].

Full-body interaction involves capturing and analyzing body movements, posture, and other physical gestures to control and manipulate digital content, opening up a world of possibilities for engaging with technology [5]. From gaming to education and art exhibitions, full-body interaction has been used in a wide range of applications to create immersive and entertaining experiences [6]–[8]. For instance, virtual reality (VR) applications incorporate full-body interaction to allow users to physically immerse themselves in a digital environment and interact with it using their entire body [9]. In educational settings, full-body interaction has been used to engage students and enhance learning throughout smart learning environments [10]. Finally, full-body interaction has also found a place in art and public installations. For example, the Rain Room exhibit by Random International allows visitors to walk through a simulated rainstorm without getting wet. The exhibition uses motion sensors and computer programming to detect the visitor's body movements and stop the rain from falling in that area, creating an immersive and interactive experience [11].

Human motion is a complex and nuanced phenomenon that involves a combination of physical movements, gestures, and body language, which makes the sensing of human motion a critical component of interaction design [12]. In full-body interaction, various technologies are used to sense and interpret the user's body movements and gestures. These technologies can include motion sensors, depth-sensing cameras, computer vision, and wearable devices. The sensing of human motion plays a vital role in enabling the computer system to respond in real time to the user's movements, which fosters a sense of flow and continuity in the interaction process [13].

1.1 Motivation

Interaction design is the process of designing interactive products, systems, and services in a way that allows users to interact with them in a significant and efficient manner. It involves the creation of interfaces that enable users to interact with technology and the designing of systems that respond appropriately to user input [14]. Interaction designers work to understand the needs and requirements of users and use that knowledge to design interfaces that are intuitive and easy to use. They also consider the context in which the interface will be used, such as the user's physical environment and the tasks that they will be performing [15]. However, both tasks require an understanding of human motion and the ability to interpret it, for which the interaction designer must resort to sophisticated algorithms and technologies to extract information from the user's bodily movements [16], [17].

Technologies to track human motion for interaction purposes can be classified into three main types: surface-based, device-based, and camera-based interaction [18]. Surface-based interaction involves using large surfaces such as walls and floors to recognize gestures. Mobile devices, touch displays, and tablets are common examples of surface-based interaction systems. Device-based interaction, on the other hand, requires users to interact with specific devices that are equipped with sensors such as pressure pads, accelerometers, and gyroscopes, to track their movement. Wearable devices, such as smartwatches and fitness trackers, fall under this category. Finally, camera-based interaction systems use optical passive or active markers, as well as color, depth, or infrared cameras, to recognize human motion. Systems that use markers are well known for being precise and expensive, being their main drawback, the requirement of wearing a tight marker bodysuit that causes discomfort [9]. As most relevant for full-body interaction, markerless systems that utilize computer vision technology are of particular interest in our research [19]. These systems rely on multiple cameras to capture and analyze human motion without the need for any special wearable add-ons or markers. This thesis specifically focuses on markerless systems because they offer several advantages. First, the use of cameras enables tracking of users and reporting of full-body information in real-time without any intrusive or uncomfortable add-ons. Second, multiple cameras in a network allow for the tracking of multiple users in large volumes without occlusions, as overlapping individual fields of view enable continuous monitoring. Finally, as computer vision technology continues to improve and become more accessible, designers and researchers can explore new and richer ways of interacting [16], [17].

Large volume spaces, such as museums, galleries, or public installations, offer unique challenges and opportunities for interaction design. These spaces provide designers with the potential to create immersive experiences for users that go beyond traditional screen-based interfaces. In such environments, full-body interaction can be particularly effective, as they offer enough room for users to move and interact with digital content without being constrained by physical boundaries. However, designing for large volume spaces also presents significant challenges, such as ensuring the system can handle multiple users simultaneously and cope with the unpredictable nature of public spaces [20], [21]. Additionally, large volume spaces require sensing technologies that are capable of tracking users accurately and robustly over a wide area [22], [23]. One example of an interactive experience that utilizes these technologies is called "Body Paint" by digital artist Memo Akten [24], [25]. The installation is set up in a large volume space such as a museum or gallery and involves multiple users interacting with the digital artwork projected on a wall.

The users are tracked by a camera-based interaction system that utilizes computer vision technology to extract features from their body movements. This combination of full-body interaction, computer vision, and large volume spaces presents a unique opportunity to create novel and compelling user experiences but requires careful consideration of the design and technological challenges involved, which is not always easy for non-technologically skilled designers to fully grasp.

It is worth noting that within the field of interaction design, there are various profiles of designers, some of whom may not possess a strong technical background and, consequently, might not fully harness the potential of motion-based feature extraction technologies [26]–[28]. This distinction is crucial and is explicitly addressed in the thesis since it represents an important factor influencing the successful implementation of such systems. By providing a clear definition of the interaction designer's profile (see Section 3.1.1), the thesis effectively illuminate why some designers may not engage with these intricate algorithms and technologies as extensively as others. Additionally, in our understanding of the interaction design field, the practitioner's work can be encapsulated by two distinct tasks, each demanding a unique set of skills and perspectives. On the one hand, designers need to determine which features to extract from users' movements and their relevance to the interaction design. This involves analyzing users' different body movements during their interaction with the artwork and selecting the ones that best communicate their intentions. On the other hand, interaction designers need to consider the user experience and ensure that the installation is intuitive and easy to use. This involves designing the interface in a way that encourages users to explore and experiment with their body movements, as well as providing feedback that guides them toward their intended interaction goals. Our argument here is that many interaction designers may not have a fully developed technical perspective to accurately interpret user movements, and complement their design-oriented thinking. Exploring these challenges and considerations in interaction design not only enriches the development of interactive experiences but also opens new avenues for the democratization of technology, making it accessible and rewarding for designers of varying backgrounds and expertise levels.

1.2 Purpose and Scope of the Thesis

Eriksson et al. [17] consider motion-based interaction as especially suited for interaction that takes place in a public or social context, providing interesting alternatives to traditional interaction techniques within social settings and public places. One of the key challenges for interaction designers is to find systematic and predictable relationships between human movement and technology, which requires sensor data analysis and extraction and interpretation of features. Far beyond very fine-grained, accurate data, motion-based interaction design often benefits from working with higher-level features to create engaging interactions. Yet, motion-based computational frameworks most often focus on the detection strategy without providing clarity on how to use such data for the design of user interactions [29]–[31]. This purely technical approach makes it challenging for non-technologically skilled practitioners to explore feature extraction technology as a design material.

Moreover, as interactive computational systems continue to advance in functionality and complexity, particularly in large volume spaces with multiple users, it becomes increasingly crucial for designers to comprehend not only the raw data provided by sensing technology but also its relevance to interaction design possibilities [12]. This thesis recognizes the importance of bridging the gap between purely technical parameters and interpretable features, serving as both a rapid prototyping and deployment tool for practitioners and a means to enhance the understanding of the diverse interaction design opportunities offered by computer vision technology. Additionally, this research acknowledges that modeling multiple-user spatial interactions is not a straightforward task if the communication strategies between users and the system are not sufficiently broad. By exploring a range of *designer-interpretable* motion-based features, we aim to expand the scope of such communication and provide guidance for designers in creating novel interactive experiences.

In light of the challenges faced by non-technologically skilled interaction designers in leveraging motion-based interaction within large-volume spaces, this thesis addresses the need for structured and predictable relationships between human motion and technology. While previous research has primarily focused on technical aspects such as sensor data analysis and detection strategies [30], [32], there is a critical gap in understanding how to effectively utilize this data for the design of user interactions.

By shifting the focus towards higher-level features and their application in creating engaging interactions, we seek to provide a comprehensive framework for motion-based interaction design. The purpose of this research is to identify full-body motion-based features that are closely tied to interaction design assumptions and generalizations, enabling the design of multi-user interactive experiences. This exploration necessitates the analysis of user-centered interaction design models, methods, and techniques, which will inform the extraction and interpretation of *designer-interpretable* features from the data produced by computer vision technology.

1.3 Working Hypothesis & Research Questions

The experimental hypothesis upon which this research is based is that *designer-interpretable* motion-based features, extracted using an easy-to-use computer vision tool, may help non-technologically skilled practitioners to design multiple-user interactive experiences in large volume spaces (up to 10x10x3 cubic meters), offering a focused research direction that investigates the correlation between users' full-body motion-based features and the opportunities for interaction design facilitated by computer vision technology. By examining this correlation, designers can gain a deeper understanding of the potential interactions that can be created in large volume spaces. This exploration opens up avenues for developing more intuitive and user-friendly interfaces, enabling non-technologically skilled practitioners to engage with interactive technologies more effectively. Under this hypothesis, interaction designers will delve into users' full-body motion-based features to establish connections between the sensing capabilities of computer vision technology and the possibilities for interaction design.

In an approach to proving this hypothesis, three independent research questions guided the process:

- **RQ1.** How have non-technologically skilled designers utilized computer vision sensing technology to design interactive spaces?
- **RQ2.** What are the key design requirements that a set of features should address when creating multiple-user motion-based interactions in the context of full-body and large volume spaces?

- **RQ3.** How can creators effectively leverage *designer-interpretable* motion-based features to produce compelling multiple-user full-body interaction experiences?

1.4 Research Aim

The objectives collectively underpin strategies designed to address the research questions and substantiate the hypothesis. In line with these intentions, this thesis aims to propose multiple-user motion-based features linked to *designer-interpretable* concepts to help non-technologically skilled practitioners create full-body interactions in large volume spaces. To achieve this aim, the following specific objectives were pursued:

- **O1.** Study computer vision's recent achievements for pose estimation and people tracking of multiple human bodies which are necessary for capturing and analyzing full-body motion-based data in large-volume interaction spaces.
- **O2.** Explore the definition of what may be considered *designer-interpretable* motion-based features, based on practitioners' interests, interaction design methodologies, and machine learning capabilities for multi-dimensional data reduction.
- **O3.** Propose a mapping strategy linking low-level motion-based features to a *designer-interpretable* feature space.
- **O4.** Evaluate the potential of a proposed set of features to become an effective tool that opens new possibilities for interaction design, artistic expression, and creative prototyping.

1.5 Research Methodology

The research methodology employed in this thesis is Action Research, a cyclical process that encompasses problem-solving, learning, and taking action to bring about transformative change and improvement in a specific field [33]. The rationale behind selecting Action Research as the underlying methodology lies in its intrinsic capacity to facilitate active researcher engagement within the domain of study. This approach enables meaningful participation with stakeholders throughout the iterative stages of Diagnosis, Planning, Intervention, and Evaluation, all of which play crucial roles in the process of problem-solving and enhancement [33]. By engaging in a dynamic interplay between theory and practice, Action Research facilitates the exploration of real-world challenges and the development of practical solutions. This proactive approach aligns seamlessly with the objectives of this thesis, which seeks to investigate and address the technical hurdles faced by designers in utilizing computer vision and raw sensor data for interactive applications. The rest of this section will provide a more detailed account of each stage, outlining the specific actions and activities undertaken during the research process.

Diagnosis. The Diagnosis stage serves as the foundation for the entire research process, encompassing a comprehensive exploration of the existing knowledge and challenges in the domain of interaction design with computer vision and raw sensor data. To begin, a rigorous literature review was conducted to gain a deep understanding of the technical state-of-the-art in the field. This involved delving into relevant academic publications, and cutting-edge research that provided insights into the current advancements and limitations. In addition to the literature review, a focus group session was organized, inviting interaction designers to participate in discussions. This interactive engagement allowed the researcher to grasp the practical tools and techniques employed by designers and their experiences with motion-based features and sensing technology. The valuable input obtained from this session contributed to identifying the prevalent challenges faced by designers in this domain.

Planning. Building on the insights gained from the Diagnosis stage, the Planning stage involved the formulation of a coherent strategy to address the identified challenges effectively. A working theory was developed, drawing from conceptual design frameworks and the designers' problem-solving approaches in dealing with technical aspects of feature representation. This theoretical framework acted as a guide, aligning the research with practical design needs in sensor-based interactions. Moreover, the Planning stage facilitated the establishment of a clear roadmap for the research, outlining the specific steps and actions to be taken in the subsequent Intervention stage. It involved devising methodologies for mapping low-level features to higher-level *designer-interpretable* concepts, providing a basis for practical implementation.

Intervention. During the Intervention stage, the research actively intervenes to address the identified challenges and implement the proposed solutions. It is a pivotal phase where the theoretical findings from the Diagnosis and Planning stages are put into action. In this thesis, the Intervention stage involved establishing a connection between low-level features extracted from computer vision and raw sensor data and higher-level features that could be easily interpreted by interaction designers. This required a thorough analysis of the technical features and their potential mapping to conceptual design frameworks. A mapping strategy was devised based on these insights to facilitate a seamless transition from technical representations to designer-friendly concepts. To validate the effectiveness of the proposed approach, a computational prototype was developed. This prototype served as a practical implementation of the research outcomes, allowing designers to interact with and tangibly evaluate the proposed features.

Evaluation. In the Evaluation stage, the research assessed the impact and efficacy of the proposed intervention. It aimed to gauge the success of the implemented changes and analyze how they addressed the initial challenges. To evaluate the intervention's effectiveness, a qualitative study was conducted, involving interviews with experts in interaction design and human-computer interaction. These interviews provided valuable insights into how the introduced features were perceived and experienced by experienced practitioners. The feedback gathered shed light on the strengths and weaknesses of the proposed solution, as well as potential areas for further improvement. The Evaluation stage also enabled a reflective analysis of the research process. The outcomes and methodologies employed were critically examined to gain valuable knowledge for future iterations on the effectiveness and practicality of the proposed design features, offering validation and contributing to the overall research outcomes.

1.6 Main Contributions

In the following list, the main contributions of this thesis are highlighted. Furthermore, they are associated with the research objectives, research questions, as well as publications during the research period (refer to Appendix A for further details).

- **C1.** *Investigating interaction designers' perceptions of using motion-based full-body features and their link to sensor-based interaction design strategies*. This contribution emphasizes the importance of understanding how interaction designers perceive and approach the utilization of motion-based full-body features in their design processes. By conducting a focus group study and gathering insights from interaction designers, the thesis seeks to shed light on their perspectives, experiences, and attitudes towards these features. The investigation delves into various aspects, such as how designers perceive the value and potential of motion-based full-body features, the challenges they face in incorporating them into their design work, and the strategies they employ to bridge the gap between sensor-based technology and interaction design principles. By examining designers' perceptions, the research aims to uncover valuable insights that can inform the development of effective design strategies and frameworks for leveraging motion-based full-body features. This contribution provides a deeper understanding of the designers' perspectives and helps identify the practical implications, opportunities, and limitations associated with incorporating these features into interaction design processes. This contribution addresses objective **O2** and research questions **RQ1** and **RQ2**, and it is discussed in detail in Chapter 3. The analysis and results were published in the *International Journal of Human-Computer Studies* [1] (see Appendix A.1 for additional information).

- **C2.** *Development of a working theory and guiding framework for aligning sensor-based interactions with the design process*. This theory emerges from the exploration of the design needs and advantages associated with the utilization of *designer-interpretable* motion-based features in multiple-user full-body interaction experiences. Through the focus group study conducted in Chapter 3, insights and perceptions of interaction designers regarding the use of motion-based features were gathered and analyzed. These findings provide a rich foundation for understanding the design needs and requirements when incorporating motion-based features in the creation of multiple-user full-body interaction experiences. The development of a working theory involves synthesizing the outcomes of the focus group study and deriving a set of principles, or guidelines, that capture the essential aspects of leveraging *designer-interpretable* motion-based features. It offers insights into how to effectively interpret and utilize motion-based data, enabling designers to create more engaging and meaningful interactions in the context of multiple users and full-body experiences. This contribution addresses objective **O3**, guided by the research question **RQ3**, and it is discussed in detail in Section 4.1. Moreover, it

[1] https://www.sciencedirect.com/journal/international-journal-of-human-computer-studies

serves as a foundational guide for the development of the tools presented and assessed in articles A.2 and A.3.

- **C3.** *Proposal of a set of designer-interpretable features and the development of a computational prototype that enables the extraction and visual representation of motion-based features*. This contribution addresses the need to bridge the gap between complex motion analysis algorithms and the practical requirements of designers. The proposed set of *designer-interpretable* features aims to provide designers with a meaningful and intuitive way to leverage human motion in their interactive experiences. These features are designed to be easily understandable and interpretable by designers, facilitating their incorporation into the design process. To support the practical implementation of these features, a computational prototype is developed. This prototype utilizes advanced algorithms and techniques to extract motion-based features from a multiple-camera setup. It then transforms these features into visual representations or descriptors that are easily comprehensible to designers. By visually representing the output of the algorithms, the prototype provides designers with a tangible and accessible way to interpret and utilize motion-based features in their design work. This contribution addresses objectives **O1** and **O3**, guided by the research question **RQ3**, and it is discussed in detail in Section 4.2. A subset of features regarding user clustering visualization was presented at the *25th International Conference on Human-computer Interaction* [2] (refer to Appendix A.2 for further details). A complete description of the set of features and the technical details involved in the development of the computational prototype are presented in an article submitted to the *ACM Transactions on Computer-Human Interaction* [3] (refer to Appendix A.3 for further details).

- **C4.** *Conducting a user-centered evaluation to assess the impact of the proposed features on concept ideation and motion-based interaction design*. This contribution focuses on gathering insights from users and designers to understand how the proposed features influence the creative process of generating ideas and designing interactions. By conducting a user-centered evaluation, the thesis aims to gain valuable feedback and perspectives from individuals who directly engage with the features in the context of concept ideation and motion-based interaction design. This evaluation involves gathering qualitative data through user interviews, and it examines how the proposed features affect the ideation phase by exploring whether they inspire new design concepts, enhance creativity, or enable novel interaction possibilities. This contribution addresses objective **O4** and is discussed in detail in Chapter 5. The results and discussion are disseminated in part in the article A.2, and in greater detail in the article A.3.

1.7 Thesis Outline

This thesis is structured to provide a detailed exploration of the concepts, methodologies, and findings related to the design and integration of motion-based full-body features in interactive experiences. The subsequent paragraphs outline the organization of this work and briefly introduce the key focus of each chapter.

[2]https://2023.hci.international/
[3]https://dl.acm.org/journal/tochi

Chapter 2, *Background and Literature Review*, summarizes the preliminary concepts and background information used in the rest of the thesis. To provide a holistic view of design frameworks for movement-based interaction, this chapter offers an overview of prominent frameworks that prioritize enhancing the user experience. By exploring these frameworks, the thesis aims to identify key principles and design considerations that can inform the development of effective full-body interactions in large volume spaces. The analysis of these design frameworks will shed light on best practices and design strategies employed by experts in the field. Furthermore, the chapter delves into the state-of-the-art techniques used for estimating and tracking the 3d pose of multiple human bodies. Accurate estimation and tracking of human movement are crucial for capturing and analyzing movement data in room-scale interaction spaces. By examining the latest advancements and approaches in this area, the thesis aims to establish a solid understanding of the technical aspects and challenges involved in capturing and processing movement data from multiple users.

Chapter 3, titled *Interaction Designers' Perceptions of Using Motion-Based Full-Body Features*, presents the results of the conducted focus group study to understand and analyze the perceptions of interaction designers toward using motion-based full-body features and how they link sensor-based interaction design with feature extraction technology. Through the analysis of the focus group discussions, this chapter sheds light on the insights, opinions, and experiences shared by interaction designers about motion-based full-body features. The study sought to explore how these designers perceive the value, usability, and potential challenges associated with integrating such features into their design processes. Additionally, it investigates how they leverage sensor-based interaction design principles and feature extraction technology to create engaging user experiences.

Chapter 4 introduces an approach that builds on the results of Chapter 3 to propose a set of *designer-interpretable* features that enhance the designer's ability to use human motion in interactive experiences. Additionally, it presents a computational prototype that extracts motion-based features and visually offers simple descriptors of the algorithms' output. The purpose of this computational prototype is to provide a practical tool that bridges the gap between the complexities of motion analysis algorithms and the practical needs of designers. Both, the proposed set of *designer-interpretable* features and the computational prototype facilitate the translation of abstract motion data into tangible design elements, enabling non-technologically skilled designers to harness the expressive potential of human motion in the creation of engaging and immersive interactive experiences.

Chapter 5, *A User-Centered Evaluation of Movement-Based Features for Interaction Design Purposes*, evaluates the effects of the proposed features on the concept ideation and design of motion-based interactions following a user-centered strategy. The chapter aims to establish the validity and effectiveness of the approach through a qualitative study involving interaction designers. It offers a detailed analysis of the designers' perspectives, experiences, and observations regarding the use of these features in their design practices. By capturing and analyzing this qualitative data, the chapter provides valuable insights into the strengths, limitations, and potential improvements of the proposed approach. Additionally, a comprehensive overview of the study design and methodology is provided, outlining the steps taken to gather data and insights from interaction designers. The chapter discusses how to refine interaction design practice by extracting motion-based features, and explores the benefits of feature visualization for designing motion-based interactions.

Chapter 6 presents a comprehensive reflection on the research conducted in this thesis, contextualizing it within the scope of the objectives and the specific purposes of the three papers derived from this study. Moreover, it concludes the thesis and presents guidelines for future work.

Finally, the complete list of publications made during the research period is presented in Appendix A.

Figure 1.1 depicts the structure of the thesis and maps the chapters to their respective objective and publications (P1 to P3).

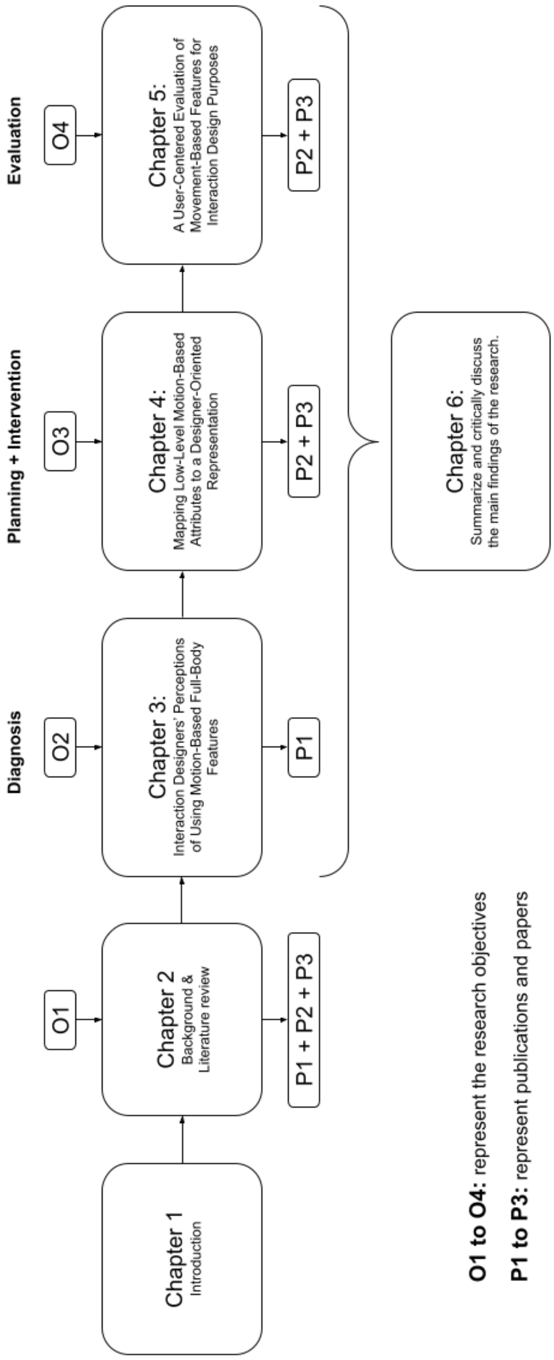


FIGURE 1.1: Thesis outline.

Chapter 2

Background and Literature Review

This chapter delves into the rich landscape of the HCI field and explores the evolution of user interfaces, focusing particularly on sensor-based interactions and full-body movement analysis. Section 2.1, sets the foundation by tracing the historical development of HCI and its impact on user interfaces. We explore the concept of a natural user interface (NUI) that emphasizes intuitive interactions, breaking away from traditional input devices. Additionally, we examine the emergence of full-body interaction, which enables users to engage with digital systems using their entire body, promoting more immersive experiences. Understanding the evolution and significance of these interaction techniques lays the groundwork for the subsequent exploration of sensor-based interactions. Section 2.2 focuses on sensor-based interactions, which form the backbone of seamless engagement between users and computational systems. By leveraging physical sensors and data capture techniques, sensor-based interactions enable users to interact with digital devices in novel and intuitive ways. This section examines the design principles and considerations involved in creating sensor-based interaction experiences, providing insights into the practical aspects of implementing such interactions. In Section 2.3, the spotlight shifts to computer vision and its key role in analyzing full-body movements. With advancements in computer vision algorithms, 3d pose estimation and tracking techniques have gained prominence, enabling accurate and real-time tracking of human movements. Additionally, movement feature extraction techniques extract meaningful patterns from the captured data, facilitating deeper analysis and understanding of human motion. This section also explores movement and feature visualization methods, which enable the representation and interpretation of complex movement data.

2.1 Human-Computer Interaction and the User Interface Evolution

HCI has emerged as a pivotal field in the realm of technology, facilitating the interaction between humans and computer systems [34]. As technology has advanced and become an integral part of our daily lives, the way we interact with computers has evolved significantly. The user interface, which serves as the primary means of communication and interaction between users and computers, has undergone a remarkable transformation over the years. Some authors have described the user interface evolution through four well-known stages [4], [35]. The first stage was the Batch Interface, back in the days of punch cards and line printers. The second stage was known as the Command Line Interface (CLI), where static instructions were used to establish a dialogue between users and computers producing a limited interaction scenario. The third stage began in the 1970s with the Graphic User Interface (GUI), derived from the "desktop metaphor" and the window, icons, menus, and pointer (WIMP) style of interaction. During the GUI stage, people started using direct manipulations, emulating real-world interactions, to control a system that responded in a predictable form [36]. Today, the Natural User Interface (NUI) stage marks an emerging technology that aims to create more intuitive, human-like communication strategies, encompassing cognitive functions like perception and expression [35].

Building upon this foundation, sections 2.1.1 and 2.1.2 will delve into two significant facets of HCI that possess the potential to reshape user experiences and interaction paradigms. First, the concept of NUI will be explored, investigating its role and implications within the broader HCI landscape. Subsequently, the focus will shift to the realm of Full-Body Interaction, a captivating avenue within HCI that harnesses the expressive and communicative capabilities inherent in the human body.

2.1.1 Natural User Interface

The term NUI is used to describe interfaces, other than a mouse and a keyboard, where the interaction is direct and consistent with our natural behavior [35]. This means, above all, that the user interaction must feel fun and easy to use based on the fact that the user may now use a broader range of basic skills compared to the more traditional GUI interaction [4]. NUIs have been shown to be successful in some specific contexts, exhibiting some common features [36]. Interactions appear in social contexts as environments for entertainment, leisure, commercial and performative demonstrations, and interactive public art. Besides, NUIs have proven to work for tasks where multiple people collaborate closely together toward a common goal. Another common characteristic is the concept of time, manifested in intermittent use and indeterminate duration which derives from playing, sharing, or social experiences [36]. Due to the sporadic and unpredictable nature of system usage, most NUIs require a seamless introduction and minimal obstacles to encourage users to become proficient. Mostly, because these interfaces are designed for voluntary users who anticipate enjoyable and fulfilling interactions.

Two major research topics in which NUIs have produced a huge impact are learning and healthcare. Bailey & Johnson [37] reviewed previous research that explores the impact of NUIs on how people learn and reported some benefits of gestures on learning and cognition. The work by Wang et al. [38] presented the design of a tangible natural user interface in the context of future classrooms to help high school students to learn the science of the human body. Authors Chatzidaki et al. [39] presented a tool that supports special educators during the assessment process of children's learning difficulties and reported that the efficiency of the usage of gestures within the game facilitates the diagnosis by offering a set of key performance indicators to the special educators. Another relevant work, that tightens together gamification and NUIs, was published by Christinaki et al. [40]. Their approach was to provide physical interaction to support early intervention and to foster facial expression learning in preschoolers with autism. In healthcare, the applications of the NUI paradigm are as prolific and diverse as for learning. Gündogdu et al. [41] argued that interaction through NUIs should further facilitate not only physical but also mental activation of people with dementia as well as stimulate sympathy and happiness. Madeira et al. [42] used NUI and gamification to encourage physical exercise in order to combat both physical and cognitive deterioration. With a similar technological approach, Rego et al. [43] explored how serious games and NUl can benefit the process of rehabilitation. To evidence the diversity of NUI applications in healthcare, the reader is encouraged to review surveys like [44], for the use of multi-touch NUIs for elders, and [45] for the applications of NUI devices for physical therapy and rehabilitation, medical operating room assistance, and fall detection and prevention.

2.1.2 Full-Body Interaction

Full-body interaction, as a concept within the HCI field, has gained significant attention due to its potential to revolutionize user experiences and interaction paradigms. Rooted in the belief that the human body possesses inherent expressive and communicative abilities, full-body interaction focuses on utilizing the entire body as an interface for interaction with digital systems [5], [46]. The theoretical foundations of full-body interaction draw upon embodied cognition, which posits that cognition is deeply influenced by the body's sensory and motor experiences [3], [34], [47]. By engaging users in physically active and expressive interactions, full-body interfaces aim to create a more intuitive and immersive connection between humans and technology [48], [49].

Within the realm of art and entertainment, full-body interaction has opened up exciting possibilities for creative expression and audience engagement. Artists and designers have embraced full-body interfaces to enable interactive installations, performances, and immersive experiences that blur the boundaries between the physical and digital domains [50]–[53]. These applications often rely on motion capture technologies and depth sensors to track users' movements and translate them into meaningful interactions. By involving the entire body, full-body interaction enables users to engage with artworks or entertainment experiences in a holistic and embodied manner, fostering a deeper sense of connection, immersion, and participation [54]–[56]. From interactive sculptures and immersive virtual reality experiences to interactive dance performances and interactive audiovisual installations, full-body interaction has become a valuable tool for artists to break new ground and challenge traditional notions of art and audience engagement.

Full-body interaction has been also used to create learning environments grounded in the embodied cognition framework [3], [34], [57]. Authors interested in this field, aim to design experiences that take advantage of the benefits of physicality to enhance the user experience and facilitate learning [52]. Malinverni & Parés [57] used full-body data to allow the inclusion of physicality in virtual learning environments, introducing the body as a mediator of meaning-making. Based on their research, the authors point out the crucial role of the body in knowledge construction in digital domains and argue consistency with cognitive science findings. Schaper & Parés [58] used a participatory design process to focus on how children can be encouraged to use their own bodies while paying attention to proxemics and to embodied constraints of the environment. Nevertheless, the authors conclude that certain body actions cannot always be directly used in the design of functional interaction experiences, leaving a gap for further research.

Some authors base their research on the assumption that full-body technologies are not always engaging for people, and that motivation is key to maintaining the user's interest [56], [59], [60]. Previous studies suggest that people usually interact with public displays or museum exhibits for less than a few minutes, demanding immediate usability strategies [61], [62]. Mishra & Cafaro [56] proposed different strategies to provide entry points to body interaction when dealing with data exploration in museums. Their results identified two strategies that worked well for this purpose: (1) Implementing multiple body movements to control the same functionality and (2) using a live representation of the users' silhouettes beside the data visualization. Muller et al. [60] employed a similar approach in examining techniques for effectively communicating interactive functionality through public displays. Their study, which drew upon findings from human visual perception research, showed that individuals tend to excel at discerning both human movements and their own reflections. The authors argued that these abilities support effective communication in such contexts.

2.2 Designing Sensor-Based Interactions

Sensor-based interaction refers to the process of interacting with digital devices through physical sensors, often employing various forms of data capture to facilitate seamless engagement between users and computational systems. These types of interfaces range from touch screens to gesture recognition, voice commands, and even brainwave monitoring. With advancements in sensor technologies and processing capabilities, sensor-based interactions are increasingly being used in the design of interactive spaces, computer-mediated environments, and user experiences [21]. Given that body movement can serve as an immediate source of input into these systems, interaction design researchers have raised questions about its potential use and consequences in human-computer interaction [12]. Motivated by recent trends towards natural interaction, controlling devices that require explicit physical contact, such as taps, buttons, and switches, are being replaced by the sensing of body movement requiring no physical contact [21]. The main effect of replacing physical objects with sensing systems is the shift in the agency of control. With sensor-based interactions, the user loses control and the system decides when to take action, maintaining operation based on the detection of changes in the environment. Ideally, this form of sensor-based interaction should be effortless, natural, and intuitive [4], [35]. However, it is often the case that users get frustrated by the new interaction paradigm in which physical cues are no longer available as the action's guide and users have to easily understand how to interact to be involved in the action [21], [60].

To guide and enhance the design process, conceptual frameworks have emerged in the interaction design field, as a set of generalizations derived from user studies, a theory, or a set of assumptions about the structure or function of phenomena [12], [16], [20]. Within HCI, frameworks are commonly used to describe a form of guidance that is explicated in a particular way to inform design and analysis [21]. Relevant frameworks that have been developed so far that offer advice on designing sensor-based interactions are presented in sections 2.2.1 through 2.2.4.

2.2.1 Making Sense of Sensing Systems

This conceptual framework borrows ideas from social sciences to inform the design of novel sensing user interfaces for computing technology [16]. Bellotti et al. present the argument of how an approach similar to that used by social scientists like Norman [63], which proposed an approximate model of seven stages of action with respect to system interaction, and Goffman [64], an interaction analyst who has written extensively on interpersonal verbal and non-verbal communication, might inform the design of novel interaction mechanisms that can be used to handle human-computer communication accomplishments [16]. As a focal point in their research, authors highlight communicative rather than cognitive aspects of interaction to present five design challenges inspired by the analysis of human-human communication: i) How does the system know I'm addressing it?, ii) How do I know what the system is attending to, iii) How does the system know what I mean when I issue a command?, iv) How do I know the system has done the correct thing?, and v) How do I recover from mistakes?

In summary, the authors highlight a distinct category of systems that avoid standard input devices like keyboards, mice, or styluses in favor of obtaining user input through sensing user actions. Their focus was on addressing the associated challenges and ensuring that each research question is thoroughly examined. By doing so, designers can effectively stay away from various potential hazards or pitfalls.

2.2.2 Expected, Sensed, and Desired

Presented by Benford et al. [65], the framework was developed to assist in the design of movable and physical interfaces and later adapted by Loke et al. [12] to focus on the movements of users instead of interfaces. The authors focused on the complex relationship between physical form and sensing technologies to help with the evaluation of how different sensing technologies match the proposed application requirements. The framework analyzes and compares the movements of the user in relation to the physical interface in terms of expected, sensed, and desired movement properties.

Expected movements are those that users naturally perform given a combination of user, interface, and environment. Along with expected movements, authors make a call to also consider unusual, although certainly possible movements as indicators that the interface is being used in an atypical manner or context. Sensed movements, as defined by authors, are those that can be measured by a computer. Therefore, they are determined by the combination of sensing technologies used with the interface [65]. Desired movements are those required for the application, and are particularly seen as relevant when an interface is used with a variety of different applications. Based on participatory and inspirational design methods, authors argue that there may be movements that are desired for the application but that are not expected and/or sensed, and other movements that are expected and/or sensed but that are not desired.

The main aspect of the authors' reflections is that expected, sensed, and desired movements only partially overlap and that mismatches between the categories can reveal not only potential problems but also opportunities. As an interesting design strategy, the authors encourage practitioners to consider the idea of compensating between all three types of movements to refine an outlined design concept toward a more detailed design specification.

2.2.3 The Sensor-Based Experience

Rogers & Muller [21] present their argument by discussing precedent conceptual frameworks proposed by Bellotti et al. [16] and Benford et al. [65]. These frameworks have predominantly offered prescriptive design advice, indicating what actions to take or avoid [21]. In contrast, Rogers & Muller focus their framework on identifying the properties of sensing technologies and the underlying user experience in sensor-based interaction. Their purpose is to outline the core dimensions of sensor-based interactions in relation to how people perceive what is going on and how this affects their understanding and subsequent behaviors [21]. In consequence, the authors argue that their framework serves as an articulatory device, helping in the definition and shaping of user experiences. Accordingly, their intention is to encourage innovative design that explores the diverse properties of sensor-based interactions, rather than focusing solely on usability design in the traditional context of human-computer interaction.

To summarize, the authors emphasize the importance of acknowledging the inherent imprecision and uncertainty of sensing systems. They argue for an alternative perspective that considers designing activities that can effectively leverage these characteristics [21]. As a specific illustration, the authors propose the incorporation of uncertainty as a deliberate and integral element within the user experience, particularly within the domain of play. They suggest that embracing and understanding uncertainty can create a stimulating user experience, and one approach to achieve this is by intentionally constructing transformations that provoke reflection [21].

2.2.4 Movement-Based Interaction in Camera Spaces

According to Eriksson et al. [17], motion-based interaction holds significant potential for social and public contexts, offering interesting alternatives to conventional interaction methods in social settings and public spaces. Cameras, being a prevalent ubiquitous sensor in motion-based interfaces, present a viable platform for innovative interfaces due to the widespread utilization of camera phones and webcams [17]. The authors' framework is built around three fundamental concepts –Space, Relations, and Feedback– which are central to the exploration of movement-based interaction through camera tracking.

The authors present Space and the physical environment as a design resource open to virtual and interactive augmentation. Using camera-based interaction authors encourage practitioners to design spaces that correspond perfectly with traditional physical spaces, where different connected but distributed spaces afford different functions and norms for social and working behaviors. A Relation describes the connection between a camera and the tracked features within the camera space. It can be described by a set of properties that defines potential interaction inputs. The number of properties depends on the algorithm used to analyze the input from the camera. The presence of a feature, the position of the feature in space, its state, identity, or information about uncertainty, are examples of properties associated with a Relation. Interaction is triggered by mapping a different action to changes in a Relation's property. For the authors, the number of relations and the number of properties associated with each relation greatly determine the complexity of the interaction. Feedback is important for movement-based interaction in camera spaces since it is considered that the interaction tool is invisible to the user [17]. Authors Eriksson et al. divide feedback from movement-based systems into input feedback and application feedback. Similar to what Bellotti et al. [16] call attention, input feedback focuses on telling the user that the input system is actually working. Moreover, similar to what Bellotti et al. call alignment [16], application feedback provides feedback about the application and its state, telling users that the system does the right thing.

2.3 Computer Vision and its Role in Full-body Movement Analysis

Computer vision plays a crucial role in full-body movement analysis by providing the means to extract, interpret, and understand human motion in a non-intrusive and automated manner. By employing computer vision techniques, researchers and practitioners can gain valuable insights into human movement, facilitating applications such as sports performance analysis, physical therapy, and gesture recognition systems. One of the primary tasks in full-body movement analysis is human pose estimation, which involves determining the joint locations and orientations of a person's body from images or videos. These pose estimation algorithms can then be used to track human motion over time, allowing for the analysis of dynamic movements and the extraction of kinematic parameters, such as joint angles and velocities. Moreover, computer vision significantly contributes to the design of motion-based interactions by enabling systems to recognize and interpret gestures, track and understand full-body movements, recognize emotions based on facial expressions and body language, and detect and classify specific actions or activities. These capabilities allow for intuitive and natural interaction with systems and devices, eliminating the need for physical controllers or touch-based interfaces. By leveraging computer vision techniques, systems can respond and adapt to user movements, leading to personalized and engaging interactions in various domains such as gaming, virtual reality, augmented reality, fitness tracking, and human-computer interaction.

In subsections 2.3.1 through 2.3.3, this thesis delves into key aspects of computer vision in motion-based interactions. First, an overview of the existing literature on techniques for estimating and tracking the 3d pose of multiple human bodies is provided, which is essential for accurately capturing and analyzing movement data in room-scale interaction spaces. The next focus is on methods employed for extracting meaningful features from movement data. This analysis showcases building upon existing work in this area and introduces innovative strategies for feature extraction. Finally, an overview of existing techniques for visualizing movement data is presented, while also highlighting the strengths and weaknesses of different approaches.

2.3.1 3d Pose Estimation & Tracking

The estimation of human pose in 3d space has been widely studied by the computer vision community, in part because of its applicability in human-computer interaction, video surveillance, and sports broadcasting. Existing approaches use a single camera view or multiple camera views to estimate either a single-person pose or multiple-person poses. In any case, the problem is divided into two stages. The first stage detects human-body landmarks in camera space (i.e., the coordinate system from the camera's point of view), and the second stage uses 2d landmark detections to reconstruct a 3-dimensional pose that matches the spatial position of people in a scene. Separating pose estimation into these two tasks allows leveraging existing 2d pose detection systems, which have achieved breakthroughs and already provide invariance to factors such as image background, lighting, clothing shape and texture, and skin color and image imperfections [66]. Single-person 3d pose estimation from a single camera view usually considers lifting detected 2d poses into 3d [66], [67] or directly regressing 3d poses [68], [69]. These methods lack reconstruction accuracy when compared to multi-view setups due to the inherited reconstruction ambiguity when only a single view is available. To alleviate this, some works exploit multi-view images effectively to obtain complementary information about the 3d scene. If multiple views are available, a 3d pose can be fully determined by simple geometry methods such as triangulation [70] based on the 2d pose in each view, or by modeling the conditional dependence between body landmarks to infer their 3d positions [71]. Some recent works [72], [73] have explored the projection of camera space features (e.g., pose heatmaps) into 3d space to regress a volumetric pose representation from the multiple-view 2d detections and estimate 3d positions of body landmarks. In the last few years, Iqbal et al. [74] proposed a weakly-supervised approach, while Xie et al. [75] suggested a meta-learning approach, both to optimize model training from multi-view setups.

As pointed out by several studies [76]–[78], multi-person 3d human pose estimation is more challenging as it should solve two key difficulties. (1) Identifying joint-to-person association in different views, and (2) handling mutual occlusions among the crowd. Early approaches used triangulation methods to create a common state space of corresponding body joints and extended the pictorial structure model to deal with multiple people [79]–[81]. Following work [82]–[84] divided the problem into two sequential sub-tasks: first, group 2d poses from different views that correspond to the same person (i.e., cross-view matching) and then, reconstruct the 3d pose from the clustered 2d poses for each person. Dong et al. [82] proposed a multi-way matching algorithm to guarantee cycle consistency across all views and a re-ID model to get appearance features for each person to enhance cross-view consistency. Kadkhodamohammadi et al. [84] computed a distance between each pair of 2d poses from different views based on the epipolar constraints and then found the cross-view correspondences with the lowest distance. Chen et al. [83] proposed to match cross-view 2d poses by applying the epipolar constraints on feet joints instead of the entire 2d pose. Zhang et al. [85] jointly formulated the temporal tracking and cross-view matching as a 4d association graph and achieved real-time performance. More recent evidence [76], [77] suggests that multi-view features may be projected into a common 3d space to avoid making decisions in each camera view. This novel approach allowed researchers to avoid the challenging cross-view matching task of previous methods, which significantly improved robustness. However, the 3d convolution used on the volumetric space is computationally expensive, thus unsuitable for real-time interactive scenarios.

2.3.2 Movement Feature Extraction

Diverse sensing technologies have made a wide range of motion data accessible to interaction designers interested in summarizing human movement to create spatial-interactive experiences, live performances, and immersive art installations [32]. Frequently used methods for recording human activity rely on motion capture systems, video cameras, or inertial measurement units, therefore feature extraction techniques must support multi-modal movement data in different representations [29]. Significant prior work has been done in geographical information science [86], expressive gesture recognition [87], and performative art studies [88] to extract information from human movement. Trajectory analysis and path comprehension are two fundamental concerns for diverse communities around geographic information science interested in the movement not only of people but of animals and vehicles [89]–[92]. In this context, movement data allow scientists to better understand the mechanisms that guide collective motion by analyzing similar trajectories of moving objects. Konzack et al. [93] analyzed the interaction between movement trajectories proposing visualizations, on a local and global scale, of delayed movement responses on simultaneously-recorded trajectories. Central to their analysis of delays, the authors present the computation of a matching between the trajectories in a so-called delay space. Ranacher & Tzavella [94] presented an overview of physical movement similarity measures in geographic information science. Authors first decompose movement into its spatial, temporal, and spatiotemporal parameters (i.e., physical quantities of movement like speed, temporal duration, or spatial path), and then review different methods for comparing movement. For a comparative analysis of trajectory similarity measures, please refer to Tao et al. [86]. Specifically, the paper compares five of the most commonly used similarity algorithms: dynamic time warping (DTW), edit distance (EDR), longest common subsequence (LCSS), discrete Fréchet distance (DFD), and Fréchet distance (FD), and present conclusions with high-level recommendations for using similarity measures in practice. However, most of the work in this area focuses on previously recorded movement data and the computational impact of algorithms and similarity measures, and not on the challenges of effectively using this data, which includes the scaling of processing systems, real-time data handling, synchronization issues [32], and the need for higher level representations of movement [95].

Preliminary work from Camurri et al. [96] presented research work concerning algorithms and computational models for real-time analysis of expressive gestures in full-body human movement. The EyesWeb processing library, by Camurri and colleagues, proposes a set of expressive cues in a layered approach to model expressive gestures from low-level physical measures up to overall motion features such as Fluency, Directness, Contraction Index, and Quantity of Motion. Following research [88], [97]–[99] advanced in the study of human movement by using similar approaches for the analysis of expressive gestures in music and dance performances. In particular, Bevilacqua et al. [99] reported on the development of a complete gestural prototype with hardware and software components, and an analysis system enabling gesture following and recognition using Hidden Markov Models (HMM). More recent work has evolved the analysis strategy to a user-centered approach based on Interactive Machine Learning (IML) and Design by Doing [100]. Presented as a promising resource to design intricate and performative movement interactions, such as embodied movement [101], IML makes it possible to design by providing examples of correct behaviors framed in terms of supervised learning. Gillies [100] points to IML as a successful method for designing movement interaction with applications in a wide range of domains: from movement-based musical interface design [102], [103] to rapid prototyping of movement in a participatory design context [104].

The movement and computing community (MOCO)[1] has shown interest in developing computational frameworks for the analysis of human movement data in recent years. Mova [29] is a movement analytics framework for motion capture data integrated with a library of feature extraction methods. The framework allows examining several of the features proposed in the literature in terms of their operative or expressive qualities. Despite that new feature extraction methods can be added, the platform lacks support for the use of input data formats more convenient for the

[1] International Conference on Movement and Computing.

implementation of real-time interactive experiences, such as video or accelerometers. OpenMoves [30] is a system for interpreting person-tracking data that emphasizes movement pattern recognition. The system was presented as a complement to OpenPTrack [105] and is therefore based on receiving people-tracking data over the local network separated into individual tracks by id. OpenMoves provides real-time centroid analysis –disregarding joint-level granularity–, low-level short-time features, and higher-level abstractions based on unsupervised and supervised machine learning techniques. Modosc [32] is a library in the form of Max abstractions that extract movement descriptors from a marker-based motion capture system in real time. The initial release of the library presented point descriptors like velocity, acceleration, jerk, and fluidity index along with descriptors to process groups of points such as center of mass, quantity of motion, contraction index, and bounding box. Dahl & Visi [32] discuss design issues that arise when working with complex marker sets regarding data handling and synchronization, and present o.dot [106] as a programming approach that handles such issues effectively. Finally, InteractML [107] is a node-base tool for designing movement interactions in Unity, based on the IML paradigm and tailored to non-experts with little programming experience. InteractML currently implements three types of ML algorithms: k-nearest-neighbor for classification, a multi-layer perceptron neural network with one hidden layer for regression, and dynamic time warping. Nevertheless, among movement features to be used as inputs to the model, the user can only choose between position, rotation, velocity, and distance to another input.

2.3.3 Movement and Feature Visualization

Alemi et al. [29] reviewed meaningful works on human movement visualization and proposed a classification in terms of their applications that we present verbatim, as we consider it appropriate. Artistic visualizations usually target the general public with aesthetic expressions and arbitrary representations. Given its abstract nature, it is far from the interest of this research and in particular of this overview. Movement summarization visualizations are used to query movement clips, compare gesture similarities, and cluster motion datasets. Finally, Analytic visualizations should provide insights into the characteristics of motion to researchers and interaction designers interested not only in movement evaluation but also in movement understanding.

A growing body of literature has investigated different ways of visualizing motion capture data [29], [31], [108], [109]. The multivariate nature of motion capture data –full of time-dependent numeric attributes–, has created a need for efficient methods for analysis that draw on machine learning, data mining, and information visualization. Frequently applied strategies derive representations from raw data usually considering the extraction of features focused on exposing and extracting as much of the semantics as possible [108]. The analytics platform by Alemi et al. [29] uses parallel visual processing capabilities of human perception to visualize multiple features at the same time and in different forms which can be used to better understand the relationships between a particular type of movement and their corresponding measurable features. A review of the literature on this topic [108] presented an overview of approaches for analyzing motion data visually. From Bernard et al.'s perspective, the characterization of the remaining challenges in the visual-interactive comparison of human motion data points to three essential factors: (1) having models that allow the extraction of features and the definition of similarity measures, (2) a user-centered approach to analyzing and comparing motion data, and (3) the need for a feedback loop with domain experts to improve the visualization tool and underlying model. Extending Bernard et al.'s work [108] to the dance and performing arts domain, Arpatzoglou et al. [109] presented a prototype of their framework DanceMoves addressing these challenges. The framework's functionality offers the interactive visual analysis of dance moves, as well as comparison, quality assessment, and visual search of dance poses. The proposed similarity measures were evaluated using agglomerative clustering over a public domain dataset and the visualization features through domain experts' feedback. However, they neglect to discuss the methodology for the qualitative study that supports their conclusions. From a different perspective, MoViz [31] is presented as a visualization tool that enables comparative evaluation of algorithms for clustering motion capture datasets. Regarding the design of behaviors and user interface, MoViz's authors made an effort to include several information visualization design principles to make the tool intuitive, informative, and accurate in data representation. Using LuminAI [110] –an interactive art installation that uses machine learning to improvise movement with human dancers– as a use case, Liu et al. [31] employed MoViz to evaluate different gesture clustering pipelines as used by the installation's AI system. As a result of this evaluation, the authors argue that the tool allowed them to identify which pipelines worked well for clustering certain datasets, which speaks to the tool's potential to understand 'black-box' algorithms better.

Chapter 3

Interaction Designers' Perceptions of Using Motion-Based Full-Body Features

Design challenges inherent in sensing systems have been of interest to interaction design researchers while looking for interfaces to actively respond to a wide variety of user behaviors (see section 2.2). From a theoretical perspective, designers have explored the relationships between movement and corresponding user-centered interaction from different angles [12], [16], [17], [21]. However, Interaction designers are constantly faced with an evolving range of technologies and usually face the common challenge of translating raw sensor data into a feature model representation suitable for its use in a creative context. To reduce the technical effort of designers to create new interaction models using the sensing capabilities that current technologies offer, it is important to understand the problems and opportunities that interaction designers perceive in using different motion-based full-body features.

In light of the previous discussion, we analyze, in this chapter, two following research questions:

- **RQ1.** How have non-technologically skilled designers utilized computer vision sensing technology to design interactive spaces?
- **RQ2.** What are the key design requirements that a set of features should address when creating multiple-user motion-based interactions in the context of full-body and large volume spaces?

To answer these questions, throughout this chapter, we examine the results of a focus group study to investigate the perspectives and attitudes of interaction designers toward the use of motion-based full-body features. The study explores how practitioners relate motion-based feature extraction technology and sensor-based interaction design methodologies, highlighting their perceptions of the conditions descriptors must meet to become a constructive exploration tool during concept ideation. Section 3.1 presents the methodology for carrying out the focus group, including the data collection and analysis strategies. Section 3.2 reveals perceptions of, and barriers to, using full-body motion-based features in a group of interaction designers to answer both research questions. Finally, Section 3.3 discuss the challenges that interaction designers often encounter in comprehending technical elements within motion-based design and explores the potential of motion-based features in understanding user behavior.

3.1 Methodology

Building upon the foundational insights provided by Oates [33], this doctoral thesis adopts the methodological framework of Action Research to explore challenges faced by practitioners in the interaction design field. As previously introduced in section 1.5, Action Research entails a cyclical process involving diagnosis, planning, action, and reflection. This structured approach provides a comprehensive means to address intricate issues and provoke meaningful transformation. In this context, the primary phase of diagnosis takes the forefront. Utilizing a qualitative method to explore the intricacies of this stage involved conducting a focus group study to gather valuable insights and perspectives. This technique acts as a potent instrument to delve into the viewpoints and attitudes of interaction designers, particularly regarding the integration of motion-based full-body features. The following subsections (3.1.1 through 3.1.4) will intricately detail the mechanics of this focus group study, covering aspects such as participant selection, methodical study preparation, procedural dynamics, and a discerning approach to data collection and analysis. Through these methodological foundations, the study aims to reveal profound insights that will contribute to the advancement of interaction design paradigms.

3.1.1 Participants

For the selection of the focus group participants, an appropriate profile with the following characteristics was defined. First, participants should have experience in interaction design from a practical perspective. We identified diverse working areas and focused our efforts on finding people with expertise in designing new interfaces, user experiences for interactive systems, new media creation, or interactive digital arts. Second, participants must have prior knowledge creating interactive full-body experiences based on movement or gestures. Potential participants were contacted with recruitment emails sent to interaction design professionals, all of them suggested by the HCI research staff of the universities to which we are affiliated. All potential participants' profiles were first reviewed to ensure that they complied with the above criteria. Eight potential participants were excluded at this stage because they didn't meet these criteria, and five further participants declined for personal reasons. In total, 12 interaction designers between the ages of 25 and 50 years (M= 34.8, SD= 8.6) agreed to participate in the focus group (see table 3.1 for details). Participants were asked for their academic background to understand professional profiles and make the most general categorization possible. To this end, we used the work by [26] as a reference. In his study about design-oriented HCI, Fallman [26] addresses diverse roles and skill sets to conceptualize different views on design and presents three competing accounts:

1. Conservative account: design is thought of as a scientific or engineering endeavor, borrowing methodology and terminology from the natural sciences, mathematics, and systems theory, drawing on a philosophical base in rationalism. A good designer in this tradition is someone who is able to follow prescribed action. Under the process-oriented conservative account, structure are at the heart of understanding and practicing design, and the view of the designer is that of an engineer.

2. Pragmatic account: rather than science or art, under the pragmatic account design takes the form of a hermeneutic process of interpretation and creation

of meaning, where designers iteratively interpret the effects of their designs on the situation at hand. The practitioner in the pragmatic account can be thought of as a reflective, know-how bricoleur, a 'self-organizing system.' The pragmatic account focuses on the situatedness of the designer in the life-world and brings to light the interweaving of roles, practices, and technologies involved in design.

3. Romantic account: it nourishes the idea of 'creative geniuses,' a legacy of the Enlightenment. Designers are seen as creative individuals with unusual talents, who often have to fight opposition to defend their unique creativity and artistic freedom. This suggests art to be a better role model for design than science. Creativity and imagination are hence seen to be the human abilities that impel design, whereas the issue of methodology is treated cautiously.

As a result, three different categories (i.e., engineer, designer, and artist) in close correspondence with the conservative, pragmatic, and romantic accounts from [26] were used to classify the study participants by role model. Our sample included 4 engineers, 5 designers, and 3 artists.

TABLE 3.1: Details of focus group participants.

Area of Expertise	View on Design (competing account)	Role Model	Technical Background	Gender
Audiovisual communication	Romantic	Artist	Non-skilled	Female
New media artist	Romantic	Artist	Non-skilled	Female
Digital arts research	Romantic	Artist	Non-skilled	Male
Creative coder	Conservative	Engineer	Strong	Male
Creative director	Conservative	Engineer	Strong	Male
Interaction developer	Conservative	Engineer	Intermediate	Male
VR content creator	Pragmatic	Designer	Non-skilled	Male
Creative director	Pragmatic	Designer	Intermediate	Male
Cognitive systems and interactive media	Conservative	Engineer	Strong	Male
User experience design	Pragmatic	Designer	Non-skilled	Male
User experience design	Pragmatic	Designer	Non-skilled	Female
Cognitive systems and interactive media	Pragmatic	Designer	Non-skilled	Male

Information systems research literature advises doing focus groups with a small number of people [111]. Groups of more than 10-12 people are difficult to moderate, and participants have little time to intervene, generating more superficial speeches

[112]. We decided to skip face-to-face discussions and hold the focus group by virtual means to take advantage of a global search for experienced participants. To manage group interaction and maintain the discussion's sense of immediacy while using a web conferencing tool, we reduced the number of simultaneous participants to a maximum of four. As [113] recommends, for conducting synchronous focus groups in cyberspace, the focus group was designed to last a maximum of an hour and a half.

Work from [114], shows that diverse groups offer better results than homogeneous groups. Diversity seems to promote richer interaction through the generation and consideration of a broader range of ideas. Reid and Reid [115] recommend forming groups with diverse profiles for research tasks, such as generating ideas. For this latter case, authors claim that online groups often outperform their face-to-face counterparts, both in the number and variety of creative ideas they produce. Nevertheless, Montoya-Weiss et al. [114] warn that communication difficulties may arise in diverse groups that must be overcome by the moderator. We decided to use a single agenda for the four different groups. Participants were assigned to groups so as to maximize group diversity on the basis of experience and profile: we included subjects with and without technical backgrounds in each group, and we made sure each group included individuals from each category (engineer, designer, and artist).

3.1.2 Study Preparation

The focus group materials were designed based on literature from both social sciences [112], [114] and information systems research [111], [116]. Three premises, closely related to the objectives of the focus group, guided the content design process:

1. Define a set of open-ended questions to obtain information on design methodologies used by participants. We rely on this freedom to invite the respondent to answer in-depth, exploring their relationship to the use of sensing technology.
2. Make use of both content and interaction data in the form of verbal and nonverbal communication aimed to understand what issues interaction designers face when trying to interpret raw data from motion-based sensors.
3. Design the focus group less than fully structured, leaving the possibility to generate questions on the spot in response to specific comments with the purpose to identify how designers relate the sensing capabilities of technology to interaction design possibilities.

We prepared a focus group guide with questions in a logical sequence to meet the research goals. During brainstorming sessions, we grouped questions and synthesized them into four main questions that divided the focus group agenda accordingly (refer to Appendix B). The first three questions corresponded to the main objectives of the focus groups, and a fourth one was aimed at driving a participative ideation stage toward the end of the session. However, another six questions were used as supplementary prompts by the moderator to lead the discussion. The ideation process was designed as a collaborative creative effort, focused on getting rich insights on appropriateness, interpretability, and other design needs for novel motion-based features they would like to use in their work. We included an ideation phase to help participants to articulate their design needs more precisely and realistically regardless of their different backgrounds. Moreover, we sought in the dialogue with the participants, ideas of solutions to the problems raised in the use of feature extraction technology.

3.1.3 Procedure

The doctoral candidate served as moderator of the focus group, previously contacting each of the selected participants to explain the objectives of the focus group, complete the informed consent form and agree on a meeting time with each group of participants. The focus group sessions were conducted using a web conferencing tool, and participants were encouraged to activate their webcams. Before beginning the focus groups, the moderator reminded participants that discussions would be used to guide the next stage of the research and that responses anonymity would be guaranteed. The first ten minutes were used for participants to introduce themselves. During this time, the moderator encouraged participants to share their professional backgrounds and experience as interaction designers. The moderator then presented the research project objectives and explained the word 'feature' in the context of this research. This was important as, given their diverse professional and technical backgrounds, the term might not mean the same for all participants.

The group discussions lasted one and a half hours, and the same moderator conducted all focus groups. All focus group sessions were recorded as video files using the web conferencing tool and later transcribed verbatim. The moderator did not take notes during the sessions; rather these were transcribed verbatim from the recordings. The full transcripts were then analyzed as detailed in section 3.1.4.

3.1.4 Data Collection and Analysis

The focus group video recordings were initially reviewed numerous times while taking notes on statements and interesting quotes, as a method of immersion and preparation stage before the analysis. Authors like [117] suggest that such an immersion process allows familiarity with the language and wording used by the participants. Data analysis was conducted by the doctoral candidate, while three other researchers reviewed the analysis criteria and validated all decisions. The research group discussed the results iteratively until a consensus was reached to maximize the objectivity of the analysis. For the focus group analysis, we followed the framework developed by Nili et al. [111], as we identified their work to be integrative and systematic. The analysis framework encompassed a systematic process consisting of seven consecutive steps, outlined in the table 3.2, guiding the progression of the analysis, with each step playing a crucial role in unraveling the intricate layers of meaning within the data. For definitions of key terms such as content area, meaning unit, category, theme, and more, please refer to Appendix C. This appendix provides concise explanations of frequently utilized concepts in qualitative analysis of focus groups, spanning various terminology used across existing literature.

Nili et al. categorize focus group data into two main groups: content and interaction data, arguing that both can be found in the form of verbal and non-verbal communication. Content data refers to any participant comment or expression that can be taken at face value and does not require knowledge of any conversation/interaction that it may be embedded in, whereas interaction data refers to agreements, questions, challenges, or support among the participants, which usually are verbal manifest communication ideas [111]. However, sometimes interaction data could also be expressed nonverbally through gestures, facial expressions, or even pitch and loudness changes in the voice. People communicate non-verbal information in social settings, intentionally or not, and such information enriches the receiver's perception of information that the encoder communicates via verbal means [118]. Thus, we considered non-verbal interaction data in the analysis.

TABLE 3.2: Steps of the focus group data analysis framework.

<table>
<tr><td colspan="2">1. Determination and organization of theoretical sensitive data types</td></tr>
<tr><td colspan="2">2. Identification of content areas</td></tr>
<tr><td rowspan="4">In each content area</td><td>3. Manifest content data analysis</td></tr>
<tr><td>4. Latent content data analysis</td></tr>
<tr><td>5. Interaction data analysis</td></tr>
<tr><td>6. Integration of results within content areas</td></tr>
<tr><td colspan="2">7. Comprehensive integration and reporting</td></tr>
</table>

Following the guideline and suggestions from Nili et al. [111], we created a complete organization scheme of theoretically sensitive data with content and interaction information. The interaction data included annotations with verbal and non-verbal communication with a low level of precision. This choice is justified by the fact that, although the focus group was designed with an emphasis on group discussion, the individual opinion is the subject of analysis. The data organization scheme considered non-verbal content and interaction data in tandem with associated verbal data because there were situations where both data types considered together, represented a better understanding of the participants' opinions. Although non-verbal data by themselves are considered meaningful by some researchers [119], in most cases gestures and facial expressions wouldn't be significant without a match to verbal communication [117].

From this data organization scheme, we identified five main content areas (see section 3.2 for further analysis and discussion), under which, all related text and non-verbal data from all sessions were merged to make the next phases of the analysis easier. First, the initial codes were examined to find any similarities and connections. Then, similar codes were grouped together into content areas that reflected the discussed topics. As the analysis continued, these content areas were refined to accurately capture the evolving insights gained from the material. The number of content areas was determined by factors such as data saturation, relevance to research questions, and the interconnections between themes. Throughout this process, an iterative and rigorous analytical approach was maintained. One can visualize it as the assembly of a jigsaw puzzle: each piece signifies a segment of the data, and as the data is reviewed and analyzed, the manner in which certain pieces interlock to create coherent images or themes becomes perceptible. These content areas, fundamentally, operate as the foundational elements for subsequent analyses, as they bring together associated data to facilitate a more focused scrutiny. At its core, the process of content area identification resembles the charting of the data's landscape, enabling the delineation of territories that would subsequently be more extensively explored in the ensuing stages of analysis. This methodical identification provides the foundation for further exploration, ensuring that the following analyses arise from a robust comprehension of the data's thematic structure.

The next phase in Nili et al.'s framework (after identifying the content areas) is to conduct a manifest analysis of content data following a bottom-up approach [111]. This phase of the analysis process consists of the following steps: (1) Identify meaning units within the manifest content of each content area. (2) Condense meaning units using a description close to their original text. (3) Label condensed meaning units with a code and sort them into categories based on similarities. (4) Express the overall interpretation of the underlying meaning for all categories in each content area via one theme. The whole process of meaning unit identification needed several iterations to highlight a phrase, sentence, or even a discussion segment that described a specific phenomenon. Going through the organization scheme of theoretically sensitive data a few times, we noticed a few themes emerging, which were developed by deductive methods. The following phase after the manifest analysis is to conduct a latent analysis of content data. In the context of the framework by [111], latent analysis is the interpretation of underlying constructs through observable elements focusing on the implied value of the data via the researcher's judgment. Similarly as before, meaning units in the latent content for each content area are first identified. Then, a description close to the content area's original text and the interpretation of each meaning unit is annotated. Finally, based on similarities among meaning unit descriptions, they are abstracted by groups into one or more themes with a corresponding label to report how data is linked with each content area [111].

The final phase in the framework is the analysis of interaction data and the definitive integration of results by content area. For the first task, verbal and non-verbal data is obtained by reading through the data organization scheme. This analysis mainly focuses on (1) identifying points of consensus or dissent with ideas expressed during discussions, and (2) interpreting the meaning of participants' interactions that indicate things other than agreement or disagreement [111]. Lastly, to capture the overall results of data analysis for each content area, all categories from the manifest analysis, and all themes from the latent analysis in each content area are merged.

3.2 Results

The process of identifying content areas involved a meticulous and systematic examination of the organized raw data, serving as a crucial bridge between the initial organization of data and the subsequent analytical stages. The primary goal here was to discern and delineate distinct sections or segments of the data that share direct connections or thematic relevance. In other words, it's about identifying clusters of information that revolve around common topics, concepts, or themes. To achieve this, the transcript, and any other forms of recorded data were carefully analyzed. Recurring patterns, phrases, key concepts, and shifts in discourse were closely attended to. As the data was progressed through, content areas where segments were intertwined by subject matter or context started to be recognized. The analysis of the organization scheme of theoretically sensitive data revealed general interaction designers' perceptions when using different motion-based full-body features. In this process, five main content areas were identified: (1) User behavior exploration, (2) Design methodologies used, (3) Relationship between interaction design and sensing technology, (4) Sensor data interpretation, and (5) Motion-based features' design needs.

The five content areas were common to all participants independently of their professional background or their previous technical experience. This section presents the content areas in the same order in which they were discussed during the focus groups and not by their importance, frequency, or uniqueness. Participants' quotes are presented to illustrate each content area. To differentiate the quotes provided by interaction design profile, those from engineers participants are denoted with an E, those from designers are denoted with a D and those from artists are denoted with an A (see sections 3.2.1 through 3.2.5). A summary of the content areas and categories is presented in Table 3.3.

TABLE 3.3: Focus groups content areas and categories.

Content Area	Category
User behavior exploration	According to the context Based on the actors involved Since conception
Design methodologies used	Good practices Methodological approaches Relationship with other peers
Relationship between interaction design and sensing technology	Technological research Best resolved at the early design stages Design away from technology Technologically influenced design
Sensor data interpretation	Valuable for design evaluation Outsourcing Tool to foster creativity
Motion-based features' design needs	Features presentation Multiple user features Single user features Criticism

3.2.1 User Behavior Exploration

Participants mentioned how they explore user behavior when developing an interactive experience in general. In this study, the exploration of user behavior consists of how designers support reflection on the experience of movement. When the design focuses on the human motion itself, a first-person perspective on the interaction design is required, closely related to the exploration of the user's movement possibilities. During the focus group, participants predominantly discussed their strategies for using the users' bodily motion during the conceptual design of an interactive experience. To a lesser extent, there was also discussion about their approaches to understanding the user's response to the experience design throughout the use of technology. Additionally, participants referred to such descriptions of the user's behavior as valuable information for understanding how to design/improve the system. Three categories emerged under this topic: (1) According to the context, (2) Based on the different actors involved, and (3) Since the conception stage.

According to the Context. Participants noted that user behavior exploration is highly contextualized by for example space, the project requirements, or the detection system technology. Concerning the use of space, participants thought that it is directly related, and that manual observations of user behavior, in situ, are required:

> E: "User behavior exploration arises from a particular need according to the use of space".

Some other participants consider it to be strongly dependent on the project target, as each project involves different dynamics:

> D: "It is very dependent on the context, and each separate project requires a different dynamic".

A group of participants with a technical background agreed upon the belief that simplifying the actions that trigger an interaction facilitates the use of sensing technologies and the extraction of motion-based features. For them, the conceptual design of an interactive experience can start from the capabilities and limitations of technology:

> E: "Starting from a certain technology capability, interaction ideas usually arise".

Based on the Actors Involved. Participants emphasized that user behavior should be explored taking into account a bigger perspective beyond the user himself. Three main actors should guide this exploration: (1) the public (users and non-users), (2) the client, and (3) the creator:

> D: "The user will have specific characteristics, but we must not forget that the client is the owner of the reason for the interactive experience, and as such, he imposes certain commercial conditions. Finally, the designer's role is to take the natural behavior of the user to enhance what is wanted from the interactive experience".
>
> E: "The starting point is always the experience objective. Based on what we need from the user, the best detection system is identified".
>
> E: "A priority is the type of audience, and something that emerges from the above is the elimination of noise from other spectators around, who are not your users".

User Behavior Explored Since the Experience Conception. Participants with an artistic professional background considered that user behavior exploration should be addressed on a multidisciplinary basis and as an integral part of the conceptual design:

> A: "We usually explore user behavior as a multidisciplinary work and from early ideation and conceptualization stages. A total fusion between creative, technical, and design processes should be sought in artistic environments".

In the discussion of the above, some of the designer background participants argued that user behavior exploration should be considered an early phase of the design process, detached from technology. However, an artist participant dissented claiming that for him, technology has always been seen as part of the discourse of artistic process:

> D: "I conceive of the interactive experience conceptualization very detached from the technical tool, in which the broad outlines of physical space and user behavior are articulated".
>
> A: "For me, the technological development of the interaction experience is a fundamental part of the artistic discourse".

3.2.2 Design Methodologies Used

Overall, participants consented that there is no general methodology that could always be followed to design interactive experiences, and most stated that it is very dependent on a multivariable context, as was shown before. Three mutually exclusive categories emerged under this topic: some participants were strictly dedicated to commenting on what they considered *good design practices*, some participants discussed their *methodological approaches*, and some others have seen in their *relationship with other peers* a design framework.

Good Design Practices. In terms of perceived good practices, some participants with technical experience thought that lowering user frustration is determinant on choosing the detection technology, for example:

> A: "The technology that reduces user frustration, that is the most determining factor in choosing detection technologies".

Some found, again, a relation with space while talking about good practices:

> E: "As a methodology, it is customary to limit the space to certain areas where interaction is defined".

Some participants think that identifying emerging technologies and being attentive to future and developing trends allows facing new interaction design challenges with the latest available technologies:

> D: "In digital arts, I usually work with things that I know that work because I have implemented them before".

Methodological Approaches. Participants were less in agreement regarding the methodological approaches to design interactive experiences. It appeared as the subtopic with clearer differences between participants with different professional backgrounds. Some technical background participants held that their methodology is based on a constant conversation between technical development and design to fit client needs, user behavior, or space requirements:

> E: "Interaction design is always a conversation between technical development and design".
>
> E: "Design Thinking is my usual methodological approach. However, it is frequent that the first stages in the methodology are cut to comply with the efficiency requirements of the client".
>
> E: "The client has an idea in his head and that is the real starting point, leaving behind creative steps such as empathize and explore".

A designer background participant showed a wider methodological perspective with three paths with different starting points (depending on the project priority):

> D: "I usually take a different approach depending on the project priorities: (1) Based on the content that must be presented to the user, (2) according to specific interaction requirements, or (3) according to the emotion you want to convey".

An artist participant shared his design methodology based on 'play and improvisation' and talked about the necessity to establish a communication/collaboration system between members of multidisciplinary design teams to define a technical-creative language:

> A: "I use play and improvisation [...], then it is important to be able to establish a collaboration system among the design team that helps define a technical-creative language".

Relationship with Other Peers. Participants with little technical experience noted that working with others is a common artistic practice. By working with technically skilled collaborators, artists can leverage their expertise to enhance the artistic process. The technical knowledge of others allows artists to explore innovative ideas, incorporate advanced technologies, and overcome technical barriers that may arise during concept ideation. This collaboration fosters a symbiotic relationship where interaction designers benefit from the technical insights of their collaborators, resulting in a more robust and impactful artistic output:

> A: "Collaborations cannot be set aside. To cover a wide spectrum of possibilities we need the support of other artistic profiles".
>
> A: "I stopped programming in my projects when I realized that there are plenty of talented developers and digital artists I can work with".

3.2.3 Relationship Between Interaction Design and Sensing Technology

For most participants, the role of the interaction designer during the ideation stages is seen as the link between members of a multi-disciplinary team who come from human sciences [120] or even graphic design [26] and computer systems programming [28]. In this context, we perceive from practitioners' discussions that in the multidisciplinary design teams they have worked with, not all members are interaction designers, nor do they have a design-oriented attitude toward movement interaction. Four categories emerged under this topic: (1) Technological research, (2) Technologically influenced design, (3) Implementation strategies are best resolved at the early design stages, and (4) Design away from technology.

Technological Research. For some participants, interaction design must be fueled by a constant evaluation of novel tools and emerging technologies to increase the ability to propose new interactions. By actively seeking out and exploring the latest advancements, designers can expand their creative possibilities and push the boundaries of what is achievable. This continuous search for new technologies allows designers to stay ahead of the curve, adapt to evolving user needs, and create innovative and engaging user experiences. Embracing new technologies not only enhances the design process but also enables designers to leverage the full potential of modern tools, platforms, and frameworks, ultimately leading to more effective and impactful design solutions:

> E: "At first, the interaction designer's experience comes into account to make decisions when making this relationship. There is also a stage of constant technology scouting to understand and manage new technologies, before being able to offer it to a client".
>
> E: "A continuous search for technical possibilities must be made to provide a solution to the client".

Technologically Influenced Design. Similar to previous technical background participants' thoughts, some designers held that technology influences interaction design. They believed that computer vision technology, despite being complex, can be utilized by non-technologically skilled designers to enhance their design process and create innovative interactions. These designers recognized the potential of motion-based feature extraction in computer vision as a valuable tool for concept ideation, enabling them to explore new possibilities and push the boundaries of interaction design:

> D: "It is impossible to be creative with something unknown. At least an influence of the visual technological culture is necessary when designing".
>
> D: "There are technological platforms (websites) that are of great help when designing. Examples like Pinterest and Vimeo allow having references and ideas of what has been done in terms of interaction design and technical development".

In that sense, technical background participants presented harder opinions, like:

> E: "You must have at least knowledge of the technological context and what is happening in terms of sensing technology, to design interactions".

Implementation Strategies are Best Resolved at the Early Design Stages. Some participants considered that taking decisions about which sensing technology to choose should be resolved at the early design stages. They emphasized the importance of considering the limitations and requirements of full-body motion-based features in order to make informed choices and avoid potential barriers during the later stages of concept ideation:

> E: "From the ideation process, you must think about the implementation to meet the design requirement. In the materialization, there must already be a decision made based on the available economic resource that determines the technological capacity".
>
> A: "Interaction design is a multidisciplinary work. Ideally, you should have several of these actors in the design process from the beginning: developers, user psychologists, and audiovisual designers".
>
> A: "Collaborative work methodologies must be built between creative and technical profiles to harmonize design and implementation from the ideation stage".

Design Away from Technology. Other participants were skeptical about giving such a big relevance to technology, in a clear minority position against most participants. They believed that excessive reliance on motion-based features could overshadow other essential aspects of interaction design, such as user experience and aesthetic considerations. Despite acknowledging the potential benefits, these designers emphasized the importance of striking a balance between technological capabilities and human-centered design principles:

> D: "Creative processes must be detached from technology, so as not to fall into doing only what the machine allows. First, you have to think about what the audience is going to feel or experience".
>
> A: "As interaction designers we must first focus on strengthening the concept, rather than choosing a sensor. I usually try to minimize dependence on technology during design, which usually results in a stronger concept".

3.2.4 Sensor Data Interpretation

There was a consensus among participants with implementation experience, arguing that knowing the nature of sensors and their output data is essential to complete an efficient development cycle. Three categories emerged under this topic: (1) Valuable for design evaluation, (2) Outsourcing, and (3) Tool to foster creativity.

Valuable for Design Evaluation. While expressing their thoughts about interpreting sensor data, participants referred to their own implementation cases for sharing ideas. Some participants argued that collecting data for evaluation of the interactive experience itself is a plus:

> D: "The possibility of having data on the behavior of users, allows us to evaluate the interactive experience design. If we think about the client's needs, it becomes a valuable product to offer".
>
> D: "The use of motion-based features can be helpful to validate whether the user reacts as expected in scenarios such as virtual reality. You can think of this as valuable information for understanding how to improve the interaction design".

For a technical developer of interactive systems, knowing both detection system nature and sensor data format is crucial to complete an effective development cycle, as the following comment suggests:

> E: "It is very important to identify the type of information that a sensor provides, to later obtain an interpretation that suits the concept of the interactive experience".

Outsourcing. Asking the participants about the sensor data interpretation revealed that some participants with technical background relate the data interpretation with outsourcing and third party software tools. Some of them prefer extensive use of SDKs and libraries to transform raw data into the required abstraction level and carefully choose a detection technology with the goal of reducing user frustration while interacting with the system. As the below comments suggest, time investment and complexity are primary concerns in the use of feature extraction technologies:

> E: "Due to time management, we do not process the raw data on our own. We are always using SDKs, libraries, and software assets to bring the raw information from the sensors to the level required by the interaction design".

> E: "We define what sensing technology to use based on reducing user frustration. We prefer to use an SDK or library to process raw data rather than develop it ourselves".

Participants expressed concern about the generalization of motion-based features' use. One participant considered that generating features for universal use in the context of interaction design is very difficult, while another held that a closed set of features will be useful only in specific contexts when linked to a certain application:

> A: "I can't imagine if a list of descriptors can be so general that they could be used in all cases. Hence, the dependency with each interaction project is shown".

Tool to Foster Creativity. Participants from all backgrounds emphasized that sensor data interpretation can be a differentiating factor and lead to creative processes that promote technological diversity even for non-technical practitioners. Participants expressed that, increasingly, there is dependence on software libraries developed by third parties for interpreting sensors and increasing production capacity. Somehow, this technical dependence is responsible for alienating interaction designers who have no coding skills but represents a tool that fosters creativity for those who do:

> E: "The interpretation of the sensor data can be a differentiator factor and lead to creative processes".
>
> A: "The experience of using ROS (Robot Operating System) may seem complex, but then the possibility of manipulating data with different programming languages, leaves the feeling that there are millions of possibilities that in the art world are not known at all".
>
> D: "In a specific case that I was faced with, studying the nature of the sensor data allowed me not only to make better use of my detection system but also to relate the capacity of the sensor to the interactive concept".

3.2.5 Motion-Based Features' Design Needs

The moderator guided the last stage of the focus group as a participative design process. We attempted to actively involve all participants in this process to help ensure the result meets their needs. We exposed participants to an initial exploratory question to focus on ideas for a solution. Through the discussion, three themes emerged: (1) Single and multiple users features, (2) Presentation of features to an interaction designer, and (3) Criticism.

Single and Multiple Users Features. Some participants made the distinction between features for single and multiple users, highlighting the need for motion-based features that can accommodate both individual and collaborative design processes. They emphasized the importance of features that enable simultaneous tracking and interpretation of movements of multiple users, facilitating group dynamics and fostering a sense of shared exploration and ideation:

> E: "Statistics of the behavior/movement of users in front of commercial spaces".

> A: "I find it interesting to have descriptors with high semantic capacity, to discriminate user actions".
>
> A: "The possibility should be left open to record certain personalized poses and make something like a matching template to recognize particular poses according to each need".

Then, in addition to this, some participants' ideas regarding features for multiple users focused on user flow, movement patterns, and grouping, recognizing the significance of capturing and analyzing interactions among users to inform the design process effectively:

> A: "I imagine features to understand movement through individual, dual and plural relationships between users".
>
> E: "All kinds of features for flow analysis, where users are grouped, circulation patterns".
>
> E: "I would like to have features to identify crossings between people". D: "Interesting having feedback on movement or posture similarities between users".

Presentation of Features to an Interaction Designer. Overall, participants showed interest in discussing how a set of features should be presented to non-technical interaction designers to facilitate its appropriation and use. Participants emphasized the importance of clear and intuitive visual representations of the features, as well as providing practical examples and demonstrations of their application in real-world design scenarios. Furthermore, they expressed a desire for interactive tools that allow designers to explore the features in real time, enabling them to gain a deeper understanding of the possibilities and limitations of each feature:

> E: "Features should be presented according to the interaction designers' ability to understand the abstraction process".
>
> D: "Many possibilities are interesting, but having a reduced view with the most popular features first, allows a faster approach to technology".
>
> E: "The set of features can be hierarchically organized by levels of abstraction, with features in at least three abstraction levels".
>
> D: "A list of possibilities allows you to identify a path to design when you are not a technology expert".
>
> D: "The semantic capacity of the feature is essential for the designer. If the tool is for designers, it is very important to use the language that the designers understand".

Criticism. A few participants were skeptical about interaction designers' overreliance on technology, and whether complex feature extraction is necessary. They questioned the need for non-technologically skilled designers to use computer vision, expressing concerns about the potential barriers it may create. Additionally, some participants expressed doubts about the practicality and relevance of utilizing full-body motion-based features, highlighting a preference for simpler and more intuitive design approaches:

> E: "As a personal conclusion and in response to the comments of the other participants, I consider that having features far beyond simple speed or acceleration would not have been useful in interactive experiences like the ones I have developed".
>
> A: "An algorithmic proposal that characterizes the user and that may serve in the design of the interaction, is not an approach that has been implemented in projects that have recently surprised me. Quite the contrary, they have come from user behavior studies or interactions observations in physical space".

3.3 Discussion

The focus group study explored the relationship between motion-based feature extraction technology and sensor-based interaction design methodologies used by practitioners, as a potential tool to define novel interaction inputs and foster user behavior exploration. Our findings extend previous research that investigated the perspectives and attitudes of interaction designers toward the use of user data in multisensory experiences [121], [122]. Previous work from Vilaza and Bardram [121] focused on designers' perspectives on shared health-data access, Seifi et al. [122] focused on novice practitioners' design needs for multimodal haptic feedback, whereas we focused on how perceptions and barriers to movement-based interaction design are formed and therefore incorporated a more practical element to understand how the practitioners' background shapes the perception of the role of technology. Our focus group sessions considered interaction designers' perceptions about how they might use interpretable features as a potential tool to explore user movement, in addition to highlighting general attitudes toward sensing technology and sensor data processing, and what might hinder or facilitate using multiple-user motion-based features.

Some content areas that emerged in this study were consistent with the literature. For example, "Technologically Influenced Design" emerged both in the current study and also in the work by Sørum [28], although the latter labeled the theme "Contribution to Products of the Future". In both studies, participants held that it would become more valuable for interaction designers to have knowledge not only in design but also in the technology field because of the increasingly interdisciplinary nature of work. In terms of design, participants in [28] who were all interaction design students, viewed it as crucial to have good knowledge of the technology to enable the creation of innovative solutions. Opinions about design teams' formation and preferences for including people with different approaches and fields of interest were also noted in both studies. A common content area with previous work by Owusu et al. [123] was the perceived impact of choosing and applying design methods. Interestingly, researchers found not only that flexible use of methods by expert designers leads to better performance, but also that the level of freedom in using methods influences novice and experienced practitioners differently [123]. Consequently, our study evidenced a consensus among participants that there is no general methodology that could always be followed to design interactive experiences, as it is perceived as a very context-dependent task. As participants mentioned, they preferred to trust their skills and common sense to guide each particular design process.

To point out how the findings of the study are useful for designers to create new interaction paradigms using the sensing capabilities that feature extraction technology offers, we first establish a relationship to the work by Sundstrom et al. [27]. The authors argue that although the mediums' properties need to be considered in any design process, technologies are usually black-boxed without much thought given to how their distinctive properties open up design possibilities. Sundstrom et al. [27] supported their study by claiming that computing technology is a more complicated material for many designers to work with [124], and we corroborated this idea. In the context of working with motion-based features, we perceived from participants that they are not very familiar with feature extraction technology as a design material which explains why only technical practitioners feel comfortable using feature extraction technologies even though they depend on tools developed by third parties. In addition, we found that participants with a design background have not paid attention to exploring and thinking imaginatively about the reach of feature extraction technologies and, consequently, they showed doubt about interaction designers' overdependence on technology. Interaction designers from all backgrounds emphasized during the focus group study, that features should provide valuable information for non-technical practitioners and help identify a path to design. To refer to both conditions during this thesis, we use the term *designer-interpretable* to describe features tailored to interaction designers. Moreover, we sought to identify defining aspects of motion-based features during the discussions with participants and then tried to find out how to reduce the technical effort from designers while experimenting with different extraction possibilities. Perspectives from participants showed that finding a design-oriented approach to present feature extraction capabilities to non-technical interaction designers should facilitate its appropriation and use. In a scenario where performing an action without implementation is fundamentally different from performing that action with feedback from technology and in the presence of tracking errors and limitations that are inevitable with movement-based technology [100], practitioners acknowledge the importance of getting a sense of technology as design materials before being used on a working prototype, especially for those that consider themselves as not technology experts.

Through the analysis of the participants' perceptions about methodological strategies for interaction design, it was clear that motion-based interaction relies on a variety of approaches that depend on the design context. Proof of this was the differences in discussing methodological approaches for designing interactive experiences. For most interaction designers with technical expertise (i.e., participants with an engineering background), technology-mediated design approaches allow incorporating a value of the moving body. We consider this perspective on interaction design in closer correspondence with "design by doing and moving" strategies [48], [100] than the more representational approaches of pragmatic-account participants (i.e., interaction designers with a design background) who design movement-based interactions detached from technology. Regarding sensor data interpretation, participants from all backgrounds coincided and argued that it is a determining factor to encourage creativity. Similarly, Hummels et al. [48] believe that designers need design methods and skills that help explore and reflect on innovative interactions and consequently present tools such as the Design Movement approach that supported and inspired novel movement-based interaction paradigms.

By asking participants how they handle technical issues while working with movement recognition technologies, we identified the importance they attach to sensor properties, such as effectiveness, interpretability, and predictability. If we relate this perspective to how participants inform design, we saw that they think it is hard to be creative with something unknown. The contribution of Sundstrom et al. [27] provides a methodology to resolve such an issue by fostering exposure to one or several of such dynamic properties considered as digital material. In other words, they advise designers to get a sense of technologies before becoming part of a working prototype. O'hara et al. [125] go deeper by drawing on the theories of embodied interaction and situated action and concern about how properties of the technology and the social system are combined in the production of meaningful and natural interaction. As pointed out by an art background participant, collaboratively work methodologies built to implement co-design strategies between creative and technical profiles in design teams may help bridge technology and material world to configure interactions in new and meaningful ways. In articulating this, we see the bodystorming scenario of the Embodied Sketching practice [126] as conclusive by encouraging designers to engage physically in co-design play-based ideation activities with peers to help sketch ideas for movement-based interactive systems.

Most participants considered that diverse and categorized motion-based features may facilitate user behavior exploration and emphasized the likelihood of using a feature extraction system with such characteristics in the future. The positive assessment of participants was further evidenced by the fact that all of them came up with ideas to meet their own design needs while in the participative design process at the end of each focus group session. Despite that, participants expressed concern about the general use of a motion-based feature closed set, knowing that interaction design is very dependent on a multivariable context. It was evidenced from participants' discussions that there is no general methodology that could always be followed to design interactive experiences and that certain sensor data descriptors will be useful only in specific contexts when linked to a certain application. Overall, participants with a background in design conceive the ideation process detached from technology. However, practitioners coming from the arts held that user behavior should be addressed on a multidisciplinary basis and considering the use of technology as part of the artistic discourse. Some were skeptical about the interaction designers' over-reliance on technology, and whether complex motion-based features are necessary to design memorable interactive experiences.

Our findings are useful to inform developers of computational tools for interactive system creation about motion-based feature refinement, thereby increasing the acceptance and adoption of higher-level features by interaction designers. Focus group participants proposed several such features, solving design needs that are worth consideration. For example, analysis of full-body movements to discriminate between multiple-users actions. Our participants did not reach a consensus in terms of how to describe multiple-user movements through feature extraction. Some participants asked for people-crossings counting, others required all kinds of features for users' flow analysis, and others showed interest in circulation patterns similarity and users' clustering in space. Another point for refinement is related to feature layout and presentation. A system for interpreting user-motion data should present features hierarchically organized by levels of abstraction and according to the interaction designers' ability to understand the extracted characteristic. Participants argued not only that a neat-presented feature set is relevant for non-technical people to favor their appropriation and use. But also that a curated list of possibilities allows practitioners to identify a path to design when they are not technology experts. Finally, from the analysis of focus groups, we argue that if we want interaction designers to understand what motion-based feature extraction technology is capable of, two things are needed: (1) developers of computational tools for interactive system creation need to work on the interpretability that designers can make of motion-based features extracted by the system, and (2) designers must recognize features' distinctive properties by experiencing them in an embodied form. By doing so, we can envision *designer-interpretable* features as a potential tool to foster user behavior exploration.

3.4 Summary

This chapter sheds light upon the effect of motion-based full-body features on interactive experience design by evidencing practitioners' approaches towards computer vision technology. With this study, we position feature extraction technology as a useful tool to better understand user behavior when designing interactive experiences, but also we acknowledge its limitations. We found that practitioners consider that processing sensor data to extract motion-based features is challenging and time-consuming. Moreover, only a few professionals with appropriate technical backgrounds feel comfortable using their own feature extraction algorithms in their interaction design work. However, most participants were eager to use a computational tool designed to interpret multiple users' motion from diverse perspectives, meaning that novel interaction strategies might emerge from a broad user-motion description capability at a reduced technical cost. Furthermore, to increase the chances of adoption, motion-based features should be grounded in conceptual frameworks known for offering advice on designing sensor-based interactions. Nevertheless, non-technical interaction designers should be aware that acknowledging the detection system's nature and being able to follow the possibilities opened up by technology is crucial to complete an effective design-development cycle. Overall, the current results are consistent with the objectives of this thesis. Based on the relationships that professionals currently make between sensor-based interaction design and feature extraction technology, it is possible to design a computational tool that reduces the technical effort of designers to characterize user movement. Finally, it has been validated from the participants' opinions that such a computational tool, conceived as a *designer-interpretable* motion-based feature extractor, constitutes a comprehensible representation of the interaction design possibilities that the sensing technology offers.

Chapter 4

Mapping Low-Level Motion-Based Attributes to a Designer-Oriented Representation

In the previous chapter, we delved into the perceptions of interaction designers regarding the utilization of motion-based full-body features. Through a comprehensive focus group study, we gained valuable insights into how designers perceive the integration of sensor-based interaction design with feature extraction technology. With this knowledge in mind, Chapter 4 takes a significant step forward by addressing research question **RQ3**. How can creators effectively leverage *designer-interpretable* motion-based features to produce compelling multiple-user full-body interaction experiences? Building upon the findings of the preceding chapter, this fourth chapter introduces a working theory that serves as the foundation for an innovative approach. Our goal within this chapter is to embark on a journey that explores the development of a framework capable of bridging the gap between low-level motion-based attributes and a designer-oriented representation.

Section 4.1 outlines the working theory that underpins our approach. This theory not only synthesizes the outcomes of the focus group study but also serves as a guide for aligning sensor-based interactions with the design process. By incorporating designers' perspectives and insights, we establish a comprehensive understanding of their specific requirements when working with motion-based features. This section provides a theoretical foundation upon which the subsequent section of this chapter is built. In Section 4.2, we present a computational prototype that exemplifies the practical implementation of our theoretical framework. This prototype is designed to extract motion-based features and offer visual representations of the algorithms' output. By providing designers with accessible and intuitive descriptors of the extracted features, we aim to facilitate their understanding and adoption of motion-based attributes within their design workflows. This section demonstrates the feasibility of our approach and offers a tangible tool for designers to experiment with and evaluate the proposed designer-oriented representation.

Ultimately, through the integration of designers' perceptions, needs, and technical considerations, we lay the foundation for the effective incorporation of human motion into interactive experiences. The subsequent sections (4.1 and 4.2) explore in-depth the details of our theoretical framework and the computational prototype, thus offering a comprehensive understanding of our proposed approach.

4.1 Working Theory

In the context of information technology research methods, a working theory is a provisional and evolving explanation that guides the design and implementation of a research project [33]. A working theory is developed through a process of observation, analysis, and reflection, and it provides a framework for organizing and interpreting data. The goal of a working theory is to propose decision alternatives and establish new relationships between theoretical perspectives. A working theory is called “working” because it changes and evolves as experiences and interactions disrupt and challenge existing ideas and assumptions. Through this process, new ways of responding and making meaning are learned. Working theories are particularly important in the field of interaction design because opinions differ between interaction designers, and so do the working theories that may guide the design process. A working theory helps designers to identify the key issues, problems, and challenges that need to be addressed, and it provides a basis for making design decisions and evaluating design solutions.

Throughout the research project, the working theory evolved and changed as we gathered new data and insights. As we conducted focus group discussions and interacted with interaction design practitioners, we gained a deeper understanding of their needs, challenges, and goals. This new information led us to refine and expand the working theory, incorporating new ideas and perspectives into our approach. Additionally, as we began to develop and test motion-based feature extraction tools, we discovered new challenges and opportunities that required us to adjust our working theory to better align with the realities of the design process. By remaining flexible and responsive to new data and insights, we were able to create a working theory that was both grounded in empirical research and adaptable to changing circumstances. As a result, our motion-based feature extraction tools were better suited to the needs of interaction designers, and we were able to make meaningful contributions to the field of interaction design research.

The theoretical framework presented here is based on the discussions and experiences with interaction design practitioners that emerged from the previous focus group study presented in Chapter 3, which revealed three main ideas:

- What interaction designers want to detect?
- What they are looking for when using features?
- How to present features to them?

Firstly, the idea of what interaction designers want to detect is crucial for the development of motion-based feature extraction tools. By understanding the types of movements and gestures that interaction designers want to detect, developers can design tools that accurately track and analyze these movements. For example, if interaction designers are interested in tracking multiple users and detecting their movement patterns, the motion-based feature extraction tools should be able to identify and distinguish between individual users and track their movements over time. Secondly, the idea of what interaction designers are looking for when using features is important because it provides insight into the specific design goals and challenges that interaction designers face. By understanding these goals and challenges, developers can design motion-based feature extraction tools that address these issues and facilitate the design process. For example, if interaction designers are struggling to reduce user frustration and simplify the use of technology, motion-based feature extraction tools could be designed to automate certain tasks or provide visual feedback to users. Thirdly, the idea of how to present features to interaction designers is important because it affects how designers use and interact with motion-based feature extraction tools. By designing tools that are presented in a way that is intuitive and easy to understand, interaction designers can more easily integrate these tools into their design process. For example, by presenting features in a hierarchical organization based on levels of abstraction, interaction designers can more easily identify which features are relevant to their design goals and how they can be used to achieve these goals.

4.1.1 What Do Interaction Designers Want to Detect?

Interaction designers are interested in detecting various aspects of user behavior, depending on whether they are dealing with single or multiple users. For single users, designers want to analyze movement in relation to space, discriminate between user actions, and estimate gestures using custom poses. These types of analyses can help designers understand how users interact with their devices and how they move around their environment. By analyzing users' movements and gestures, designers can gain insights into how to design interfaces and interactions that are more intuitive and natural.

For multiple users, interaction designers are interested in identifying relationships among users, detecting people flow and grouping, and analyzing movement patterns. These types of analyses can help designers understand how users interact with each other and how they move around shared spaces. By analyzing users' movements and relationships, designers can gain insights into how to design interfaces and interactions that are more socially aware and inclusive.

Designers can use a range of tools and technologies to detect these aspects of user behavior. For example, they can use cameras, motion sensors, or other tracking technologies to analyze users' movements and gestures. They can also use software to analyze data collected from these sensors, generating visualizations and other types of feedback that can inform the design process.

4.1.2 What are Designers Looking for When Using Features?

Interaction designers need movement-based features to design interactive systems using computer vision technology. We believe it is possible to generalize by saying that to define such features, we must start with purely technical necessities and achieve significant representations close to interaction design. In discussions with practitioners, we have identified requirements that gradually address the needs between one extreme and the other: simplify the use of technology, eliminate noise from non-users detections, reduce user frustration, evaluate user behavior, validate user actions, link sensor capabilities with design concepts, foster creativity, and open new possibilities for interaction design. These requirements highlight the importance of considering both technical capabilities and user needs when designing interactive systems based on movement-based features. By prioritizing these factors, designers can develop more effective and efficient solutions that meet the demands of all stakeholders involved.

First, designers need to understand the technical capabilities of the sensors and other technologies they are using. This understanding can help them identify potential limitations or opportunities that may affect the design process. Once designers understand the technical capabilities of their tools, they can start to explore how to use those tools to create more intuitive and user-friendly interactions. One way to achieve these goals is to start by simplifying the use of technology. This means removing unnecessary complexity and focusing on the core features and functions that are most essential to the user experience. Designers can also work to eliminate noise from non-users' detections, such as by filtering out irrelevant or unwanted data.

Next, designers need to focus on reducing user frustration. This means designing interfaces and interactions that are easy to use and understand, even for users who may be unfamiliar with the technology. Designers can also evaluate user behavior to gain insights into how users are interacting with the system and identify areas for improvement. To validate user actions, designers need to ensure that the system is providing feedback that is both meaningful and relevant to the user's actions. This can help users understand the system's capabilities and limitations, which can in turn lead to more successful interactions.

Linking sensor capabilities and design concepts is another important goal for interaction designers. By understanding the technical capabilities of their sensors, designers can identify new opportunities for interaction and create more innovative and engaging designs. Fostering creativity is also an important goal for interaction designers. This involves investigating innovative applications of sensors and other technologies, enabling designers to push the boundaries of possibility and craft interactive experiences that are not only immersive but also remarkably compelling. Finally, interaction designers need to be open to new possibilities for interaction design. This means exploring new technologies, experimenting with new approaches, and embracing new ideas and perspectives.

In order to better understand what interaction designers are looking for when using motion-based feature extraction tools, we developed a visual summary of our findings (see Figure 4.1). This summary reflects how the needs of interaction designers depart from a technical perspective, such as simplifying the use of technology and reducing user frustration, and consistently end up as pure interaction design needs, such as fostering creativity and opening new possibilities for interaction design. As essential reference, based on our analysis of previous focus groups and experience with interaction design practitioners, this visual summary informs our working theory.

4.1.3 How to Present Features to Interaction Designers?

When it comes to presenting features and capabilities to interaction designers, it is crucial to strike a balance between ease of understanding and relevance to their specific needs. Interaction designers play a vital role in crafting intuitive and engaging user experiences, and their ability to grasp and leverage the capabilities of technology is key to their success. However, due to the diverse backgrounds and expertise levels of interaction designers, presenting features in a way that resonates with them can be a challenge.

To foster effective communication, it is essential to adopt strategies that cater to the designers' varying levels of technical expertise, simplify the presentation of features, utilize designers' language, organize features hierarchically, and inspire

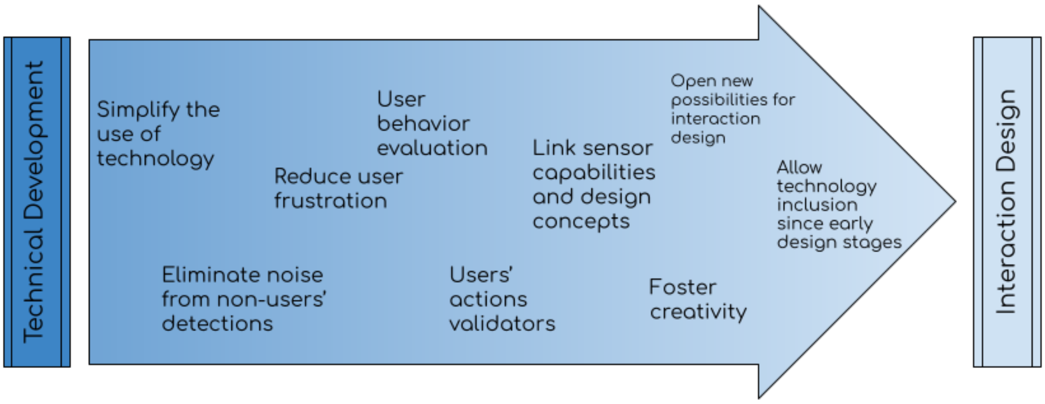


FIGURE 4.1: Visual summary of interaction designer needs when using motion-based feature extraction tools.

innovative thinking. By tailoring the presentation approach to their specific requirements, interaction designers can more seamlessly integrate technology into their design process and create compelling interactive experiences. In the context of our discussions with practitioners about motion-based features' design needs during focus groups, we will now delve into a range of strategies. These strategies are explicitly inspired by these discussions and can be employed to present features and capabilities to interaction designers. Their purpose is to enhance designers' understanding, align with their design goals, and stimulate their creativity. By adopting these strategies, designers are empowered to effectively harness the potential of technology, resulting in the delivery of seamless, intuitive, and delightful user experiences.

Tailoring the Abstraction Process to Designers' Understanding. One of the key considerations when presenting features and capabilities to interaction designers is their varying levels of technical expertise and understanding. Designers come from diverse backgrounds, some with extensive technical knowledge, while others may possess limited programming skills. To ensure effective communication, it is essential to present features in a way that aligns with each designer's level of knowledge and experience. This may involve providing clear explanations, visual aids, or demonstrations that help simplify complex concepts. By tailoring the abstraction process to designers' understanding, they can quickly grasp the functionalities and make informed design decisions.

Gradual Introduction of Features to Common Technological Approaches. When introducing features, it is beneficial to adopt a gradual approach that starts with the most basic and commonly used functionalities. This allows designers to familiarize themselves with the core capabilities before delving into more advanced features. By presenting features in a simplified and organized manner, designers can develop a solid foundation of knowledge and build confidence in utilizing technology to enhance their designs. Additionally, providing real-world examples and practical use cases can help designers relate the features to their own projects, making the learning process more engaging and applicable.

Semantic Description in Designers' Language. Communication between interaction designers and technologists can sometimes be hindered by the use of technical terminology. To bridge this gap, features and capabilities should be described using the language and terminology familiar to interaction designers. This means avoiding excessive technical jargon and instead using terms that align with the designers' domain. By employing designers' language, the presentation becomes more accessible, and designers can easily comprehend and articulate their design requirements, fostering a collaborative and productive environment.

Hierarchical Organization for Understanding Relationships. Presenting features and capabilities in a hierarchical manner can greatly assist interaction designers in understanding the relationships and dependencies between different functionalities. By organizing features into a logical structure, designers can navigate the complexity more effectively. This hierarchical organization helps designers identify the core building blocks of the technology and how they can be combined to achieve desired outcomes. Understanding the relationships between features enables designers to make informed decisions about the sequencing and integration of functionalities, resulting in more cohesive and coherent interactive experiences.

Presenting the Full Range of Possibilities for Innovation. To inspire innovation and creativity, designers should be exposed to the full range of available features and capabilities, even if they may not be immediately relevant to their current projects. By showcasing the breadth of possibilities, designers are encouraged to think beyond the constraints of their immediate design needs. This exposure enables them to identify novel applications, unique interactions, and innovative design solutions. By exploring the full potential of technology, designers can push the boundaries of their creativity, leading to the development of more compelling and engaging interactive experiences.

By adopting these effective strategies and presenting features in a manner that is tailored to interaction designers' needs, designers can seamlessly integrate technology into their design process. This approach empowers them to leverage the capabilities of technology effectively, resulting in the creation of interactive experiences that captivate and engage users.

4.2 Implementation of a Computational Prototype

The practical implementation of our theoretical framework is demonstrated through the development of a computational prototype. This prototype serves as a tangible manifestation of our conceptual groundwork and is designed to extract motion-based features, subsequently providing visual representations of the algorithms' outputs. These visual descriptors of extracted features offer a level of accessibility and intuitiveness that aims to significantly enhance interaction designers' comprehension and incorporation of motion-based attributes into their design workflows.

Foremost, the idea of discerning what precisely interaction designers aspire to detect holds paramount importance in shaping the development of motion-based feature extraction tools. This understanding is pivotal as it guides the design of tools capable of accurately tracking and analyzing the specific types of movements and gestures that designers aim to use. For instance, if interaction designers seek to trace multiple users' distinct movement patterns, the motion-based feature extraction tools must be adept at distinguishing individual users and monitoring their motions across time. Subsequently, the idea of how to effectively present features to interaction designers is also important, influencing how these professionals interact with and exploit motion-based feature extraction tools. The design of tools must be a harmony of intuitiveness and simplicity, facilitating seamless integration within designers' creative processes. For instance, structuring features hierarchically based on levels of abstraction empowers interaction designers to readily identify relevant features aligned with their design objectives, effectively guiding their application in achieving these goals.

Through the implementation of the computational prototype, not only is the feasibility of our approach showcased, but interaction designers are also provided with a tangible toolset for experimentation and evaluation, thus bringing the comprehensive implications of the theoretical framework to realization in the practical domain of interaction design. Grounded in this rationale, we created an extraction and visualization prototype that comprises three primary components: (1) a 3d pose estimation and tracking component, (2) a feature extraction system, and (3) the visualization engine.

4.2.1 Pose Estimation and Tracking in 3d Space

The 3d pose estimation and tracking component considers a multi-camera approach to account for large volume spaces and resolve occlusions and limited field-of-view problems. The procedure for this component consists of a sequence of steps: First, a calibration step based on using a *ChArUco* board pattern [127] to capture properly synchronized images, and then obtain camera poses by bundle adjustment optimization. Second, a people detection and pose estimation step based on running the top-down *AlphaPose* estimator on each view. Third, a multi-view 3d pose reconstruction step. And fourth, a tracking and filtering step to finally obtain smoothed human poses of all the people in the scene. Further elaboration of the last three steps is warranted, as these crucial components enable accurate and robust human pose estimations in challenging environments and are therefore essential to the success of the overall system.

4.2.1.1 People Detection and Pose Estimation on Each View

The first step in estimating 3d poses from a set of multiple calibrated cameras is to detect 2d poses on each view. To this end, we adopted the body-only AlphaPose estimator [128] which follows a top-down strategy. Alphapose is a two-step framework that first detects human bounding boxes and then independently estimates the pose within each box. Furthermore, AlphaPose's architecture is optimized for easy usage and further development as various human detectors and pose estimators can be used for custom purposes. In the current implementation, an off-the-shelf YOLOV3 pre-trained detector [129] and an efficient high-accuracy pose estimator named FastPose [128] are adopted. FastPose network structure uses ResNet152 [130] as the backbone to extract features, three Dense Upsampling Convolution modules for feature up-sampling, and a 1x1 convolution layer to generate heatmaps. To speed up the process, each view is pre-processed using AlphaPose before the multi-view matching step. Estimated 2d poses in COCO [131] format, bounding boxes, and cropped heatmaps are saved as they are needed for the next step.

4.2.1.2 3d Pose Reconstruction

To match the estimated 2d poses across views, we need to find the bounding boxes belonging to the same person in all views. To solve this problem we adopted the approach of Dong et al. [82] in which appearance similarity and geometric compatibility cues are combined to calculate the affinity score between bounding boxes. In this context, appearance similarity is a measure of the distance between appearance descriptors of a bounding box pair obtained from a person re-identification network (i.e., the cropped image of each bounding box is fed through the network to extract a feature vector for each bounding box and then the Euclidian distance is computed between the descriptors of a bounding box pair). Furthermore, geometric compatibility means that corresponding 2d joint locations should satisfy the epipolar constraint (i.e., a joint in the first view should lie on the epipolar line associated with its correspondence in the second view).

The muti-view correspondences are calculated from the affinity scores and represented as a partial permutation matrix that maximizes the corresponding affinities and is also cycle-consistent across multiple views (i.e. any two corresponding bounding boxes in two views should correspond to the same bounding box in another view). The bounding boxes with no matches in other views are regarded as false detections and discarded. The multi-way matching algorithm groups 2d poses of the same person in different views from which 3d poses can be reconstructed. To incorporate the structural prior on human skeletons, Dong et al.'s algorithm makes use of the 3DPS [80] model in a reduced state space for efficient inference.

4.2.1.3 Tracking and Filtering

For 3d pose tracking, we implemented a tracking-by-detection method to find correspondences between adjacent frames. A major challenge in our approach is how to associate unreliable detection results with existing tracks. To this end, we first collect redundant candidates from outputs of both detection and tracking and then use a unified scoring function and non-maximal suppression for optimal selection [132]. Later, to improve the identification ability we compute the appearance feature similarity between non-redundant candidates and existing tracks. This similarity function was also used in the construction of the affinity score in the cross-view correspondences and in this case, featured as a key component in data association to address the spatial distance limitation. Given the combined similarity, the correspondences can be solved with the Hungarian algorithm.

Missing joints are caused by noise detections or occlusions and usually degrade the temporal consistency. To remove outliers and to infer missing joints a simple effective smoothing algorithm was adopted [133]. Temporal averaging is used to fill in missing joints, while a Gaussian kernel with standard deviation σ is used to smooth each joint trajectory. We explored different values of σ to visually understand the effects of smoothing and adjusted the value empirically. As Tanke & Gall [133] suggested, higher values of σ improve the performance when noise 3d poses are given at the cost of reducing accuracy.

4.2.2 Movement-Based Features

The proposed features for motion-based interaction design are motivated by the observation that interaction designers are not very familiar with feature extraction technology as a design material, which limits their ability to explore and think imaginatively about its possibilities [95]. The previous study (refer to Chapter 3) identified general needs and limitations that interaction designers face when using feature extraction technology, and evidenced that these requirements needed to be translated into specific *designer-interpretable* features. To bridge this gap, the prototype provides a concrete set of features that meet the demands of interaction designers when designing motion-based interactions. The proposed features aim to reduce the technical effort required from designers while experimenting with different extraction possibilities, facilitate the appropriation and use of feature extraction capabilities by non-technical interaction designers, and provide tools that can help determine a course of action for the design process when using motion analysis algorithms. Subsections 4.2.2.1 through 4.2.2.7, explain each feature algorithmically and provide evidence from previous work that validates their effectiveness and relevance in practical design scenarios. The selection of seven features is driven by a deliberate rationale. The initial three features delve into characterizing individual movements, laying the foundation for capturing nuanced gestures and actions. Complementing these are the remaining four features, which are rooted in the description of group behaviors formed through spatial clustering or trajectory similarity among individuals. This comprehensive range is intended to holistically address both individual and group dynamics within motion-based interactions.

In summary, the decision to incorporate seven features was guided by the intent to encompass a diverse spectrum of movement aspects and interaction scenarios. The hierarchical organization of features by increasing complexity further demonstrates the systematic approach taken to empower designers in creating meaningful and engaging motion-based interactions. It is important to emphasize that these choices have been justified based on the identified gaps in the field and the goal of facilitating more accessible and imaginative use of feature extraction technology by interaction designers.

4.2.2.1 Bounding Box

At each time step, we calculate the rectangular parallelepiped enclosing the 3d pose of each person. The bounding box is fairly straightforward and strictly speaking, is not even a motion-based feature. However, the decision to provide it reflects our intention of exploring how interaction designers interpret motion from even the simplest feature. We expect that from the bounding box variation over time, the interaction designer may abstract useful representations of the user's motion. As an example of the above, Glowinski et al. [134] reported having used bounding shapes in the analysis of affective body gestures. Figure 4.2(a) shows two examples of the bounding box feature visualization overlaid on a rendered 3d view of the scene.

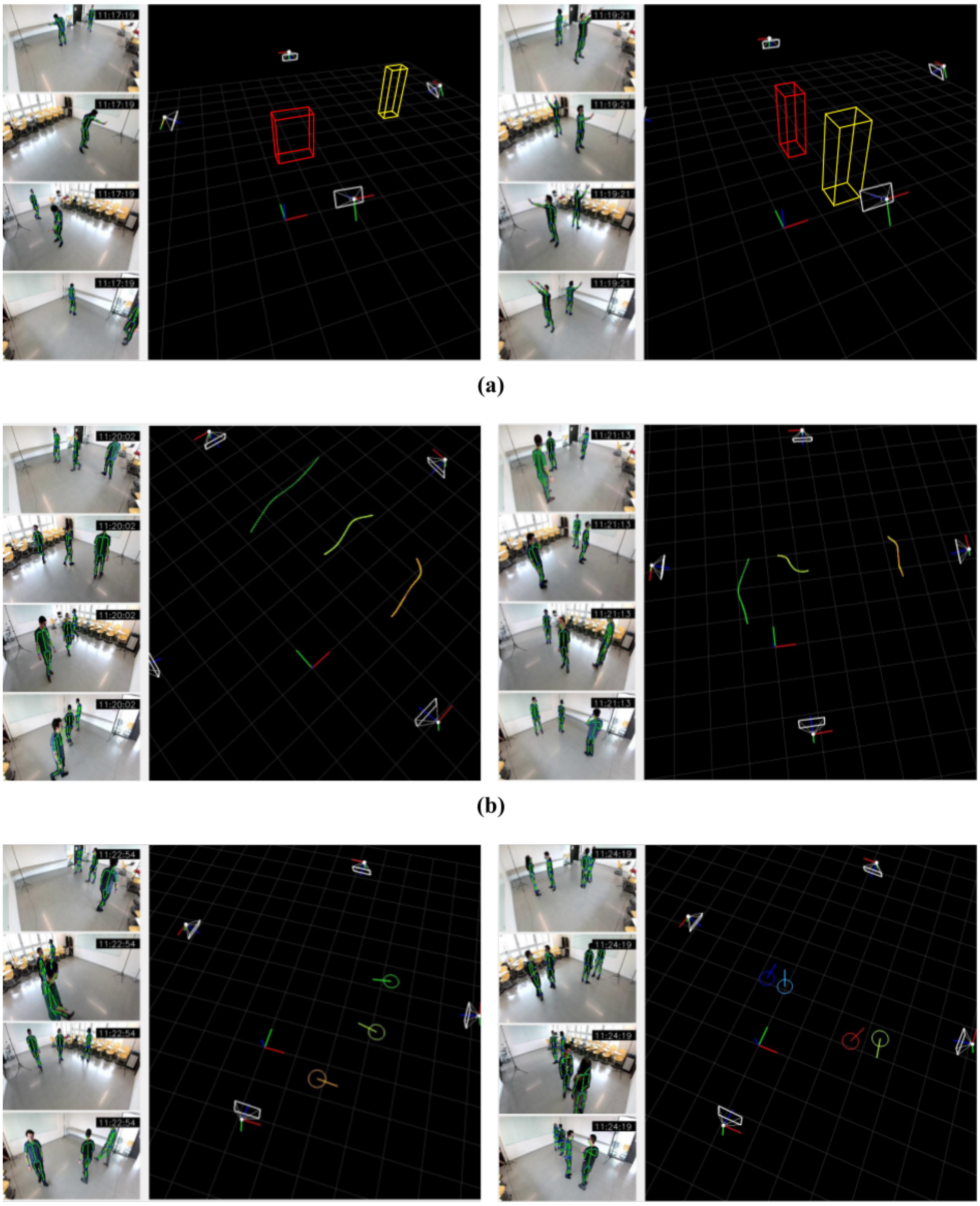


FIGURE 4.2: Visualizations of individual movement features. (a) bounding box, (b) trajectory, and (c) heading.

4.2.2.2 Trajectory

We present the user trajectory as a short-time feature as it characterizes recent movement by considering a fixed number of the most recent position points. User trajectories are updated into a circular buffer of three-dimensional points at every time step, therefore, a trajectory represents a sequence of points in 3d space. By default, the positional data for the trajectory representation is the central point between the shoulder joints, however, any other body joint could be used. A new trajectory object is created when a new user appears in the scene, storing id and relevant data from the user tracking module. The user trajectory feature also serves as an input for group behavior features based on time-windowed positional data, as will be discussed in section 4.2.2.6. Refer to figure 4.2(b) to see examples of the trajectory feature visualization overlaid on a rendered 3d view of the scene.

4.2.2.3 Heading

We approximate the linear velocity vector of the user trajectory using the difference between the last two position points. We discard the z-dimension as we are only interested in visualizing the trajectory heading over the ground plane. We assume that users are moving forward and express their heading as the angle between the linear velocity vector and the x-axis. Using this angle, we simply plot a fixed-size arrow from the current user position. Figure 4.2(c) shows examples of the heading feature visualization overlaid on a rendered 3d view of the scene.

4.2.2.4 Instantaneous Clustering

The instantaneous clustering feature takes the set of user centroid positions at the current time step and finds clusters using the mean shift algorithm [135]. Particularly, unsupervised learning is attractive to the context of this study, given the capability to characterize group behavior without reliance on a priori knowledge. The mean shift algorithm is widely used in data analysis because it's non-parametric and doesn't require any predefined number of clusters. As Amin & Burke [30] pointed out, the mean shift algorithm fits well with the frequently changing nature of social interaction scenes and live performances. In an effort to produce a meaningful feature visualization, we decided to represent the distance of users to the cluster's centroid to which they belong with lines on the ground plane, and the number of users in each cluster as the radius of a circle representing the cluster. Figure 4.3(a) shows the instantaneous clustering feature visualization overlaid on a rendered 3d view of the scene.

4.2.2.5 Hotspots

The hotspots feature provides the capability to identify frequently-visited areas or routes in the space. Based on the work of Amin & Burke [30], hotspots are addressed as a long-term, macroscopic form of clustering. Nevertheless, we decided to perform clustering to positional data over a fixed time window using a different algorithm. We have chosen the DBSCAN algorithm as it performs density-based clustering that is robust to outliers. DBSCAN works on the assumption that clusters are dense regions in space separated by regions of lower density [136], which provided better hotspot visualization results when compared to other clustering algorithms. Regarding the feature visualization, the prototype shows simple descriptors of the algorithm's output such as position, spread, and boundaries of clusters in a heatmap representation to enrich the user's perception of group behavior. Refer to figure 4.3(b) to see examples of the hotspots feature visualization overlaid on a rendered 3d view of the scene.

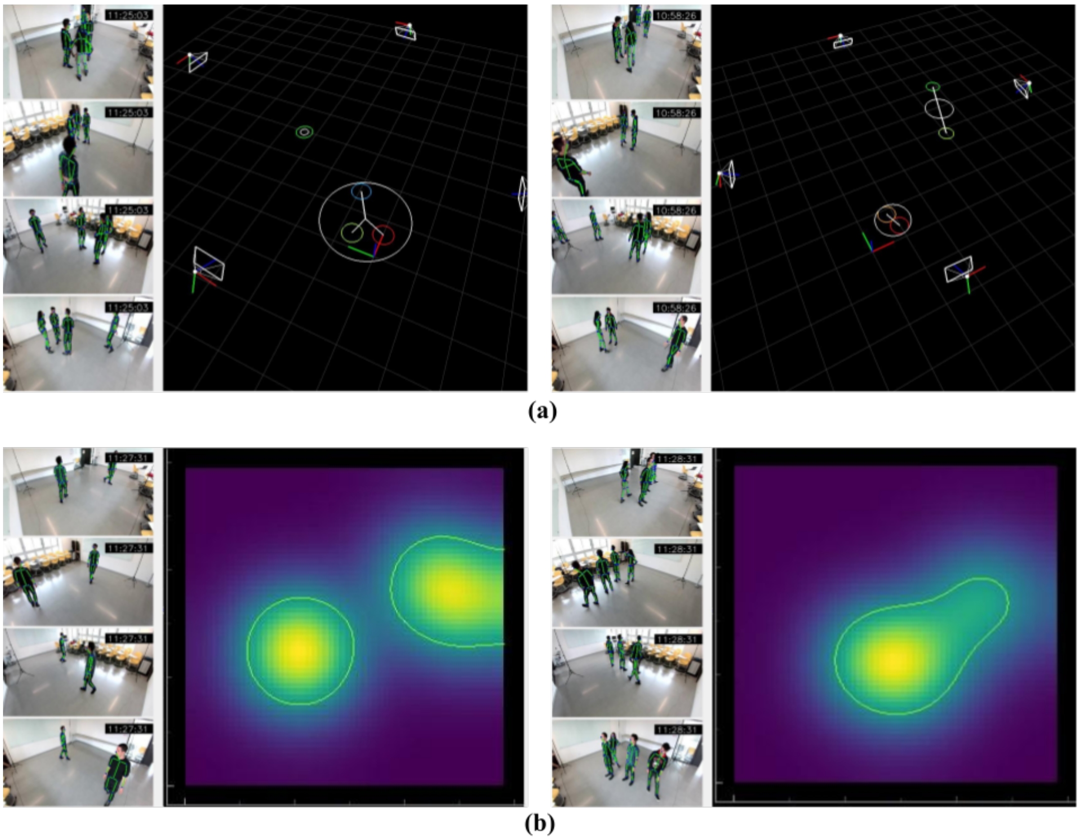


FIGURE 4.3: Visualizations of user clustering features. (a) instantaneous clustering, and (b) hotspots.

4.2.2.6 Trajectory Similarity

The trajectory similarity feature accounts for the detection of instantaneous common patterns of movement based on all user trajectories. Given that user trajectories are of the same size (see section 4.2.2.2), we perform a Procrustes analysis [137], [138] between all trajectory pairs to obtain a similarity matrix from which a mean score is obtained. In the context of the Procrustes analysis, an optimal transformation is first applied to one of the trajectories to obtain a scale, rotation, and reflection invariant similarity measure, meaning that the similarity score is independent of user position. For a proper visualization, we format the feature as an event-based variable that outputs a 'positive' detection if the mean score is above a threshold, meaning that all users in the scene were moving similarly. Refer to figure 4.4(a) to see examples of the trajectory similarity feature visualization overlaid on a rendered 3d view of the scene.

4.2.2.7 Correlations Across Movement Patterns

Derived from the trajectory similarity feature, we examine the similarity matrix to detect actors moving similarly. Instead of a detection event, the prototype graphically indicates whenever users correspond with each other by performing a synchronized movement. With a feature like this, we provide interaction designers with the ability to draw comparisons between users in the scene and to better understand inter-user dynamics. Figure 4.4(b) shows this feature visualization overlaid on a rendered 3d view of the scene.

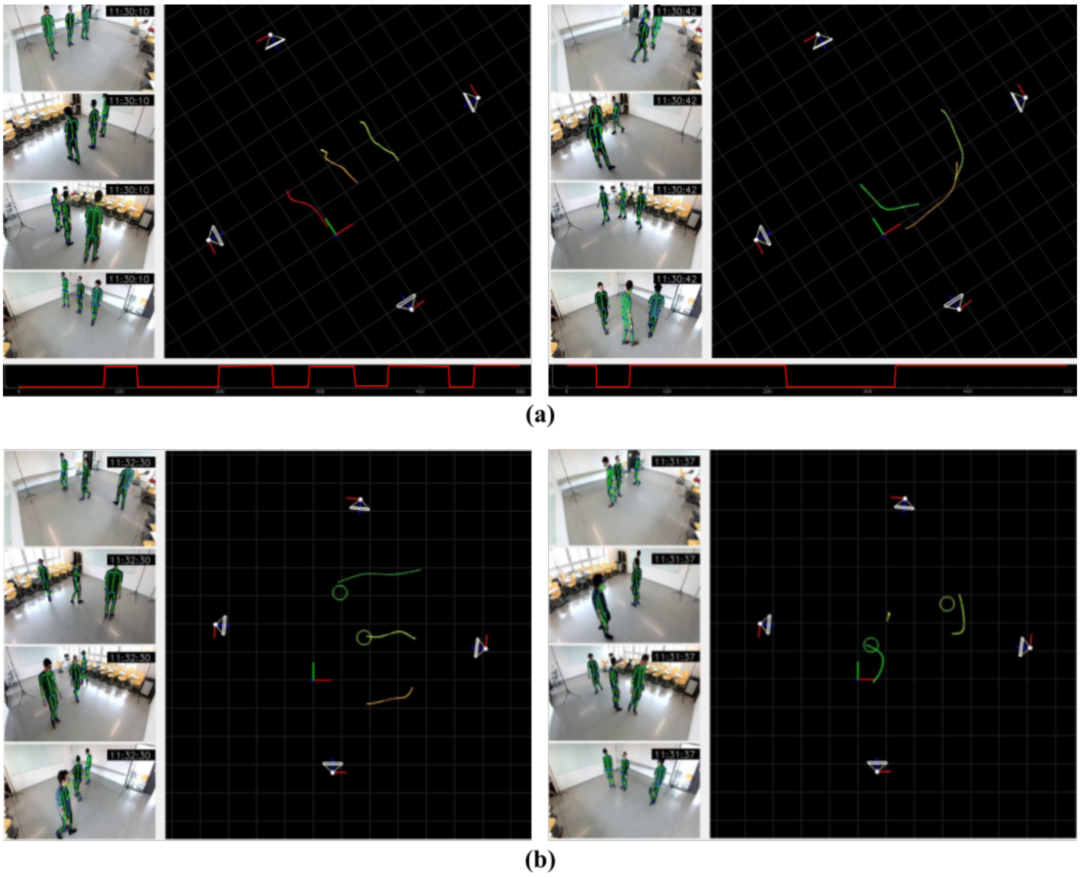


FIGURE 4.4: Visualizations of group behavior features. (a) trajectory similarity feature, showing user trajectories and event-based detection at the bottom of the window, and (b) correlations across movement patterns, with circles indicating instantaneous position of users performing synchronized movements.

4.2.3 Visualization Engine

The visualization engine presents a 3d view of the scene that includes visual representations of the human poses and extracted features. Figure 4.5 displays the graphical user interface, which is composed of three sections. The selection section contains controls for opening scenes and choosing features, while the multi-view video playback section overlays 2d pose detection results on each view. Additionally, a three-dimensional representation of the scene is included, where the feature visualization is rendered. The prototype user can control the camera position in the 3d scene using six degrees of motion input via the keyboard keys and mouse. This enables the user to explore the relationship between features and space from various perspectives, navigate the 3d scene, and decide whether to focus on a smaller region or zoom outwards for a broad overview. To minimize the user's focus on technical intricacies related to algorithm tuning, we have opted to limit the level of intervention required to use the prototype. This decision was based on observations from previous work [95] that found less technical designers lacked the ability to address algorithm tuning, and providing them with access to this parameterization could be confusing, deterring them from using the results on creation and conceptualization stages. As a result, users are not able to engage in algorithm parameterization or receive numerical feedback. By removing these technical barriers, we aim to make the prototype more accessible and user-friendly, allowing a wider range of interaction designers to take advantage of its results without getting bogged down in the technical details of algorithm tuning. In addition, the prototype adheres to Tufte's conception of graphical integrity in which visual representations of data should neither overrepresent nor underrepresent its effects and phenomena [139].

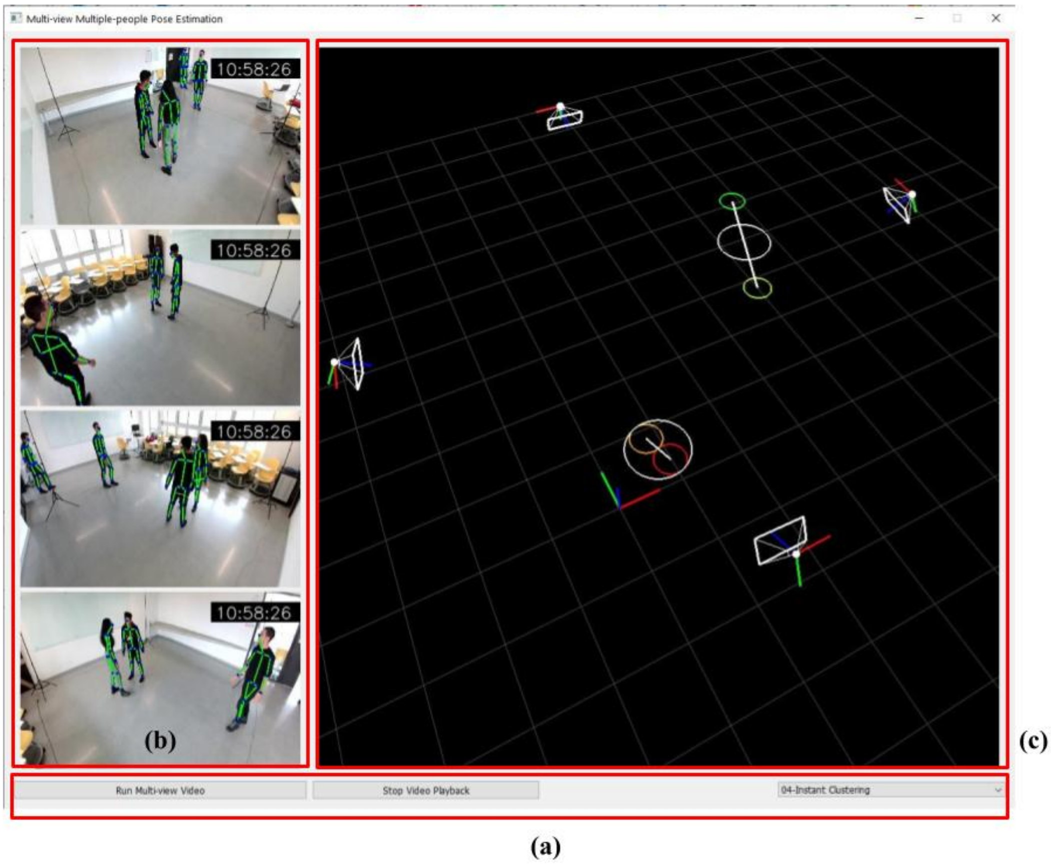


FIGURE 4.5: Graphical user interface of the feature extraction and visualization prototype. (a) selection section, (b) multi-view video playback section, and (c) three-dimensional representation of the scene and feature visualization section.

Chapter 5

A User-Centered Evaluation of Movement-Based Features for Interaction Design Purposes

The growing interest of the human-computer interaction community in exploring the motion-based interaction design space has been approached from two different perspectives. One notable exploration involves movement models like Laban Movement Analysis' Effort qualities, which have sparked discussions on perception, experience, and the nuances of human motion [140]–[144]. These approaches, however, reveal limitations in terms of demanding user expertise and concealing subtleties [145]. In parallel, computational frameworks have arisen, illuminating human movement intricacies through feature extraction [29], [30], [32], [99]. Nevertheless, little is understood about how such movement characteristics are perceived by interaction designers or experienced by end users. Hence, we consider it paramount to understand if designers identify functional, performative, and expressive aspects in the features provided by computational frameworks.

This chapter evaluates the impact of proposed features on interaction design, aligning with objective **O4**. A user-centered evaluation is undertaken, with qualitative insights gathered directly from designers. By exploring their experiences, the goal is to determine if novel design concepts are inspired, creativity is fostered, and new interaction possibilities are introduced. In alignment with **RQ3**, the investigation bridges theoretical and computational perspectives. The intent is to understand how designers harness these computational features and how end users experience their impact. Ultimately, this chapter aims to determine whether the proposed features effectively stimulate innovative interaction design possibilities, cultivating creativity, artistic expression, and avenues for inventive prototyping.

The work presented in this chapter builds on the focus group study (see Chapter 3), which conveys interaction designers' insight that mapping purely technical parameters to a set of *designer-interpretable* features constitutes an intelligible representation of the interaction design possibilities offered by sensing technology. In order to further examine this idea, this chapter describes a qualitative study that involved interviews with six interaction designers. Specifically, the study analyzed the impact of motion-based features on the design of spatial interactions by presenting multi-view recorded scenes to the interviewees. By comparing design strategies with and without the support of the computational prototype presented in Chapter 4, the study makes the following contributions:

1. Supplies evidence that feature visualization facilitates the design of movement-based interactions by allowing designers to use a graphical language to explore

the relationship between individual movement intentions, grouping patterns, and space.

2. Highlights the importance of the perceived agency by end-users as a driving factor in practitioners' decisions when incorporating movement-based features into their interaction design practice.

For the rest of the chapter, Section 5.1 explains the methodology for carrying out the interviews with interaction designers, including the data collection and analysis strategies. Additionally, Section 5.2 presents the results of the qualitative evaluation. Finally, Section 5.3 examines how motion-based feature extraction can refine interaction design practice, and explores the benefits of feature visualization.

5.1 Study Design

The primary goal of the study was to investigate whether the proposed features might be helpful for interaction designers working in room-scale interaction spaces and how they can incorporate them into their own work. Since the investigation follow a user-centered approach, a qualitative study was considered to develop an understanding of interaction designers' needs and perceptions. A semi-structured interview-based methodology was chosen to ensure the coverage of the breadth of experiences practitioners have when designing interactive experiences and to observe the impact of movement-based features and their visualization in design practice.

5.1.1 Participants

This qualitative study involved six interaction designers. To ensure a diverse perspective, both familiar and new participants were included. Specifically, three participants had been part of a previous study (refer to Chapter 3), where a focus group methodology was employed. The remaining three participants were selected exclusively for this new study and had not taken part in any prior research related to the perception of motion-based features. The decision to include both sets of participants was deliberate and rooted in the research methodology. Individuals with a professional profile that aligned with the criteria of the previous study were sought to complete the group of participants. Specifically, practitioners with experience in diverse areas of interaction design, such as creating new interfaces, shaping user experiences for interactive systems, engaging in new media creation, or contributing to interactive digital arts were considered. Furthermore, participants were required to possess prior expertise in developing interactive full-body experiences based on movement or gestures.

The strategic mix of participants allowed for the capture of fresh insights from new interaction designers while maintaining continuity by including individuals who had contributed in the previous study. It is worth noting that the shift in participant numbers compared to the focus group study was a deliberate choice. In this study, individual interviews were employed rather than group discussions. This approach enables a deep dive into each participant's perspectives without the cross-information exchange and ideation dynamics that characterize focus group settings. The age range of the participants spanned from 23 to 45 years (M= 30.3, SD= 9.6), and their professional experience varied from 3 to 15 years (M=7, SD=4.6). Geographically, two participants were based in Spain, two in Colombia, and two in the United States, providing a diverse range of insights and experiences.

Inclusion of Both Sets of Participants. By involving participants who were part of the previous focus group study, continuity is established. These individuals have already contributed valuable ideas and insights during the ideation process in the earlier study. Their participation in the current study provides insight into their thoughts and reactions regarding the computational prototype based on their prior input. This not only enriches current findings but also creates a sense of investment and ownership among these participants. On the other hand, including new participants brings in fresh perspectives. These individuals have not been exposed to the dynamics of the previous study, which ensures that their reactions and ideas are not influenced by prior group interactions. Their inputs can provide a novel angle and potentially challenge or validate the ideas generated in the earlier stages of research.

Shift in Participant Numbers. The decision to involve fewer participants in this study compared to the previous focus group approach is influenced by the change in methodology. Employing individual interviews as opposed to group discussions allows for in-depth exploration of each participant's thoughts without the influence or dynamics of group interactions [14]. By conducting one-on-one interviews, one can extract nuanced insights, allowing participants to express their opinions without the potential conformity or negotiation that might occur in a group setting. The individual interview approach allows the participants' thoughts, reactions and perceptions to be explored in greater detail. Each interview becomes an opportunity for a deep dive into their experiences, preferences, and critiques. The qualitative richness of the data collected through individual interviews can compensate for the smaller participant number, as the depth of insights obtained can be substantial.

5.1.2 Study Preparation

A room-scale interaction space was set up with a total area of approximately 5m x 4m. The space was equipped with four high-definition cameras placed at various positions, always above the height of people, to capture the movements from different perspectives. A multi-camera configuration was designed to record each view and enable the tracking of multiple people in a space with occlusions and limited field-of-view problems.

A set of seven social interaction scenes, tailored for each proposed motion-based feature, was designed to produce the stimuli for the study. These scenes depicted four actors engaged in interactions within motion-based settings, and their recording resulted in video clips lasting two minutes each. No behavioral constraints were imposed on the subjects except for a contextual scene description. This approach facilitated the emergence of unscripted, authentic interactions, eliminating contrivances in their movements. The actors were provided with a strategic scene description, enabling them to customize their performances in alignment with one of the seven proposed motion-based features (see section 4.2.2). This alignment ensured that the actors' performances seamlessly interfaced with the computational prototype, facilitating the subsequent extraction of rich insights from their natural interactions. Three of the seven scenes explore the intricate characterization of individual movements, highlighting the nuanced essence of each actor's gestures. The other four scenes delve into collective behaviors through spatial clustering and trajectory patterns. In the context of motion-based interaction, the scene descriptions provided a foundation while allowing actors to naturally respond to their fellow performers' movements and interpretations. For detailed information on the scene descriptions and other relevant instructions given to actors, please refer to the Appendix F. A multiple-view frame of each recorded scene is shown in Figure 5.1.

In addition, each multi-view recording was processed using the 3d pose estimation and tracking component, and 15 body landmark locations of each actor in 3d space were extracted. An example of the multi-camera tracking component capturing the movements of the actors in the interaction space can be seen in Figure 5.2. Finally, stimulus pairs were prepared from each recorded scene. Condition A presents the multi-view recording with no additional information, while condition B has the people's pose overlaid on a 3d-rendered view of the scene. This latter arrangement will subsequently serve as input to the feature extraction system, allowing the prototype's user to select a feature from a list for visual rendering during playback. As they are easier to interpret, during the study sessions, the individual movement stimuli were first introduced, followed by the group behavior stimuli. This allowed participants to gain confidence in their understanding of the features and to slowly build up their own vocabulary. Figures 5.3 and 5.4 provide a visual representation of the stimulus pairs used in the study, which helps readers better understand the methodology and results presented in section 5.2.

5.1.3 Interview Method

Before conducting interviews, the doctoral candidate contacted selected participants (identified as P1 through P6), explained the study's purpose, completed informed consent forms, and agreed on meeting times. Interviews were conducted using a web conferencing tool, with all participants activating their webcams. The interviewer reminded participants that discussions would be used for research purposes only, and that responses would be kept anonymous. Finally, the interviewer presented the research aims and explained the contributions of the previous study. Participants in the interview were introduced to the prototype's graphical user interface and encouraged to interact with it through the interviewer's actions, with the option to request functions like stop and resume playback during the study.

Participants engaged in a semi-structured interview designed to play the role of a pre- and post-evaluation of participants' insights, allowing us to ascertain the utility of feature extraction and visualization in enhancing their analytical capabilities. In the pre-evaluation stage, participants were exposed to the initial stimulus–a multi-view video capturing a scene without additional contextual information. They were prompted to evaluate whether people's movement within the scene held significance for interaction design. Participants were encouraged to envision spatial-interactive experiences that could be cultivated based solely on the visual cues provided by the video. This phase aimed to establish a baseline understanding of participants' expectations, prior experiences, and attitudes toward the technology. The subsequent post-evaluation stage introduced participants to the complementary stimulus in the pair. Here, participants reexamined the same scene, equipped with visual support from the feature extraction and visualization prototype. The objective was to explore the impact of feature extraction on their interpretation of people's motion and to stimulate innovative interaction design ideation.

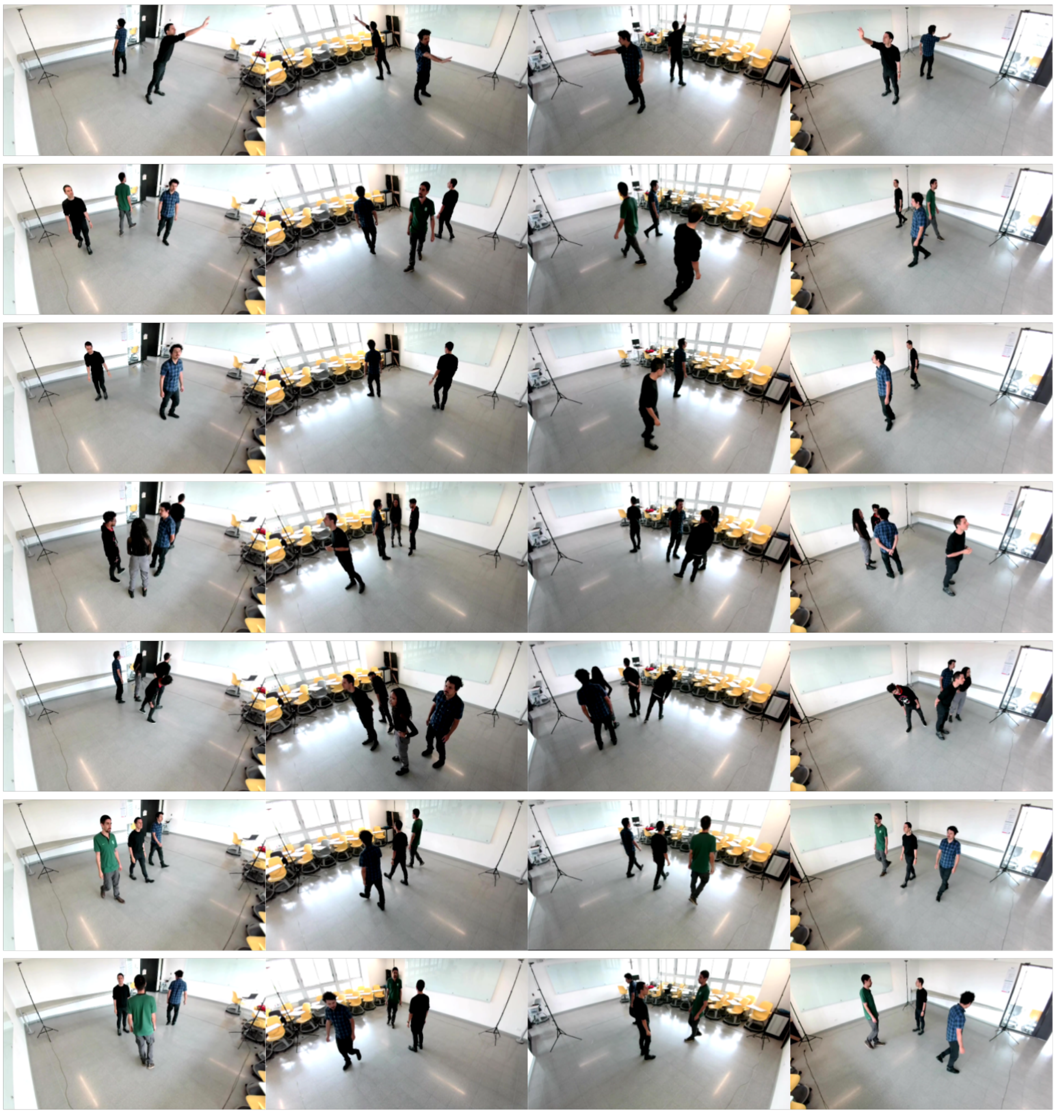


FIGURE 5.1: Multiple views of interaction scenes used to produce the stimuli for the study. Each row represent a different scene. The first three scenes target individual movements of the actors, while the remaining four target group behavior.

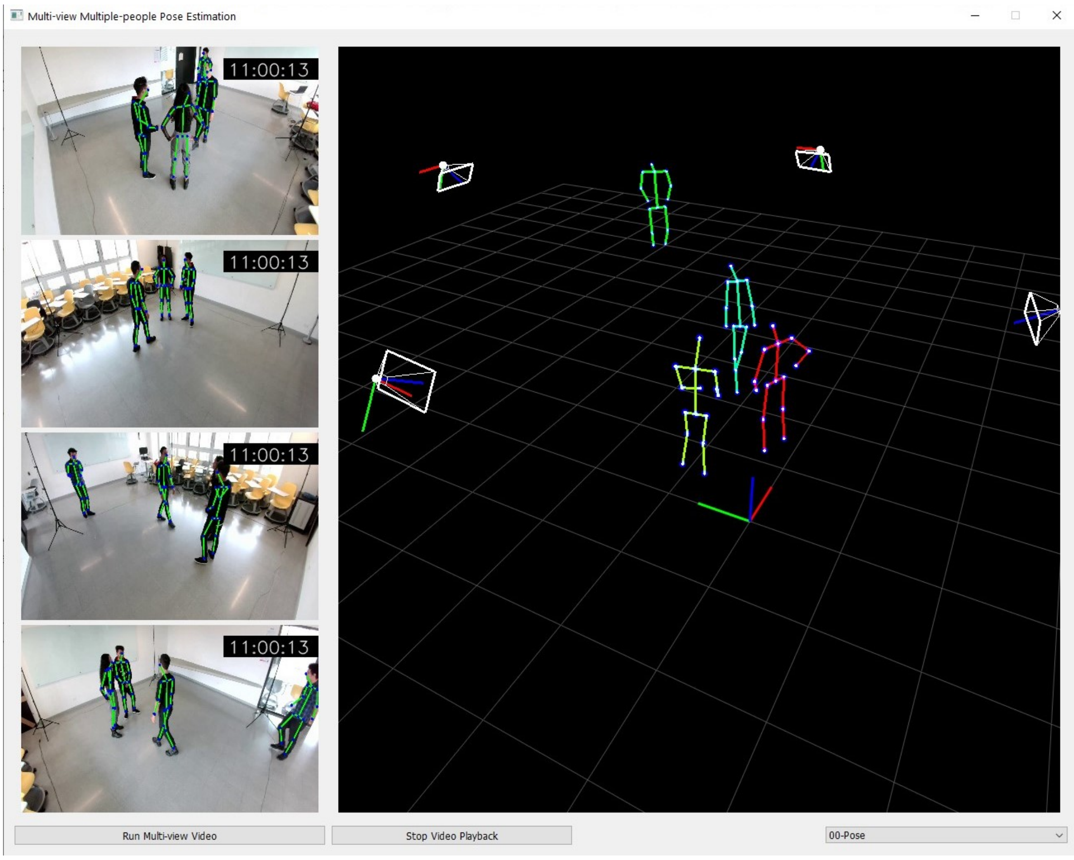
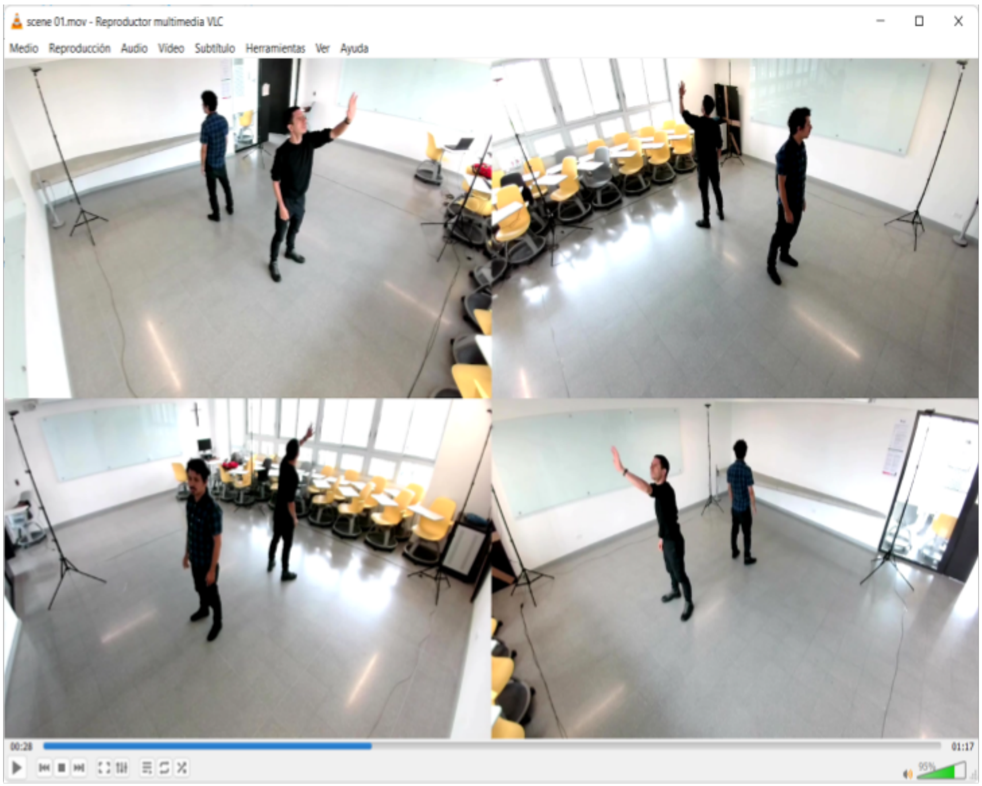


FIGURE 5.2: Multi-view multiple-people tracking system.

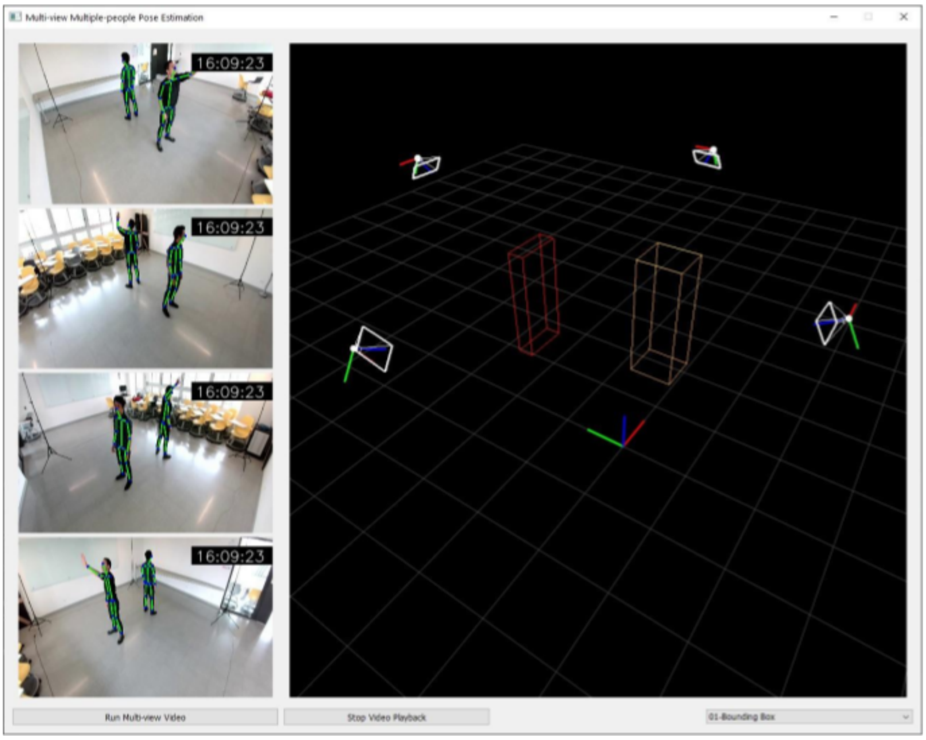
The process of analyzing stimuli and generating insights was iterated across all seven stimulus pairs. Participants were afforded the opportunity to pursue topics of interest that organically emerged from the discussions, fostering a comprehensive exploration of various motion-based interactions. As the interviews drew to a close, participants were invited to reflect on their holistic experience. They were queried about factors that influenced their engagement, the challenges associated with assessing people's movement without proper visual cues, and the perceived limitations of the process. Refer to Appendix D to see the complete interview protocol.

5.2 Analysis and Results

A total of 656 minutes were collected in the form of 6 interviews, ranging from 90 to 135 minutes in duration (M= 109.3, SD= 17). All sessions were recorded using a web application that facilitated collaborative annotation of noteworthy moments for later review. Initially, the interview recordings were reviewed, with key insights highlighted. Turning points and relevant segments were identified using timestamps, and these insights were subsequently revisited and examined by the research team. Excluding introductory remarks by the interviewer and off-topic discussions, the corresponding audio recordings were transcribed. Grounded Theory methods were used for qualitative analysis to develop theories or concepts based on the data itself rather than starting with preconceived theories or hypotheses[14]. The textual information was segmented into discrete statements and categorized by condition

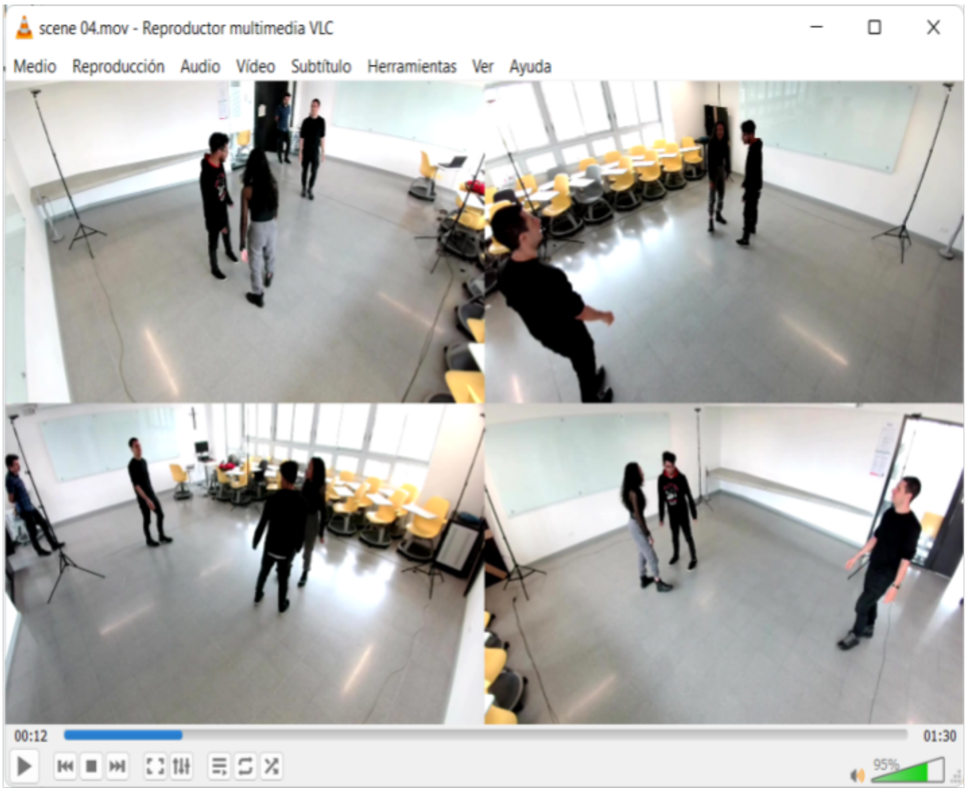


(a)

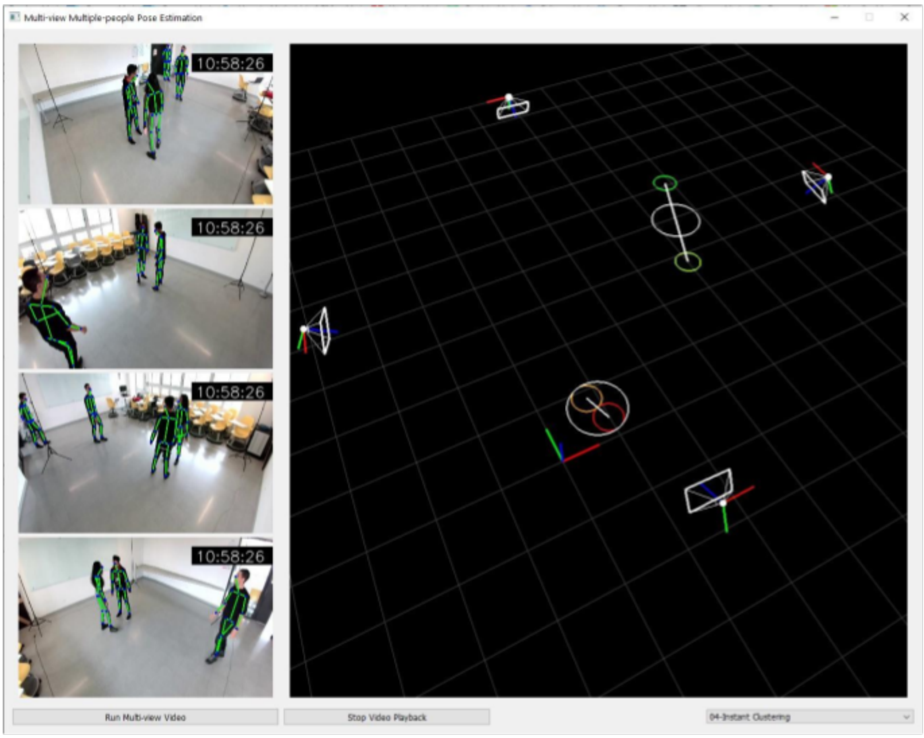


(b)

FIGURE 5.3: Stimulus pair used for discussing the 'Bounding Box' feature. (a) Condition A: multi-view recording with no additional information. (b) Condition B: same scene with visual support from the feature extraction and visualization prototype.



(a)



(b)

FIGURE 5.4: Stimulus pair used for discussing the 'Instantaneous Clustering' feature. (a) Condition A: multi-view recording with no additional information. (b) Condition B: same scene with visual support from the feature extraction and visualization prototype.

rather than participant, fostering an inductive approach that prioritizes the participants' contributions to generating analytical themes [116]. To aid in comprehending the analysis process, appendix E provides definitions of Grounded Theory concepts that have been employed throughout the qualitative analysis.

The analysis began with open coding, which involved breaking down interview transcripts into smaller units and assigning descriptive labels or codes to each unit. This initial coding phase facilitated an exploration of the raw data, enabling the emergence of initial concepts and the identification of categories as they naturally evolved. The process of refining and revising codes and categories was ensured by continuously comparing newly coded data with previously coded segments, to accurately represent the data. The number of categories that emerged was determined by the principles of data saturation, relevance to the research question, clarity, and ongoing review. Next, the axial coding phase helped to further refine and deepen the qualitative analysis by building on the initial category identification achieved during the open coding phase. The purpose of axial coding was to better understand the connections between categories, the contextual factors affecting them, and their importance in clarifying the focal phenomenon being examined. Through axial coding, a richer and more nuanced understanding of the research subject emerged, revealing the underlying structures and processes that shaped the data. This iterative approach allowed the derivation of a core set of central themes from the interview data. The resulting coded statements and themes were discussed by the research team over several rounds, and decisions were made by consensus.

The rest of this section reports on the findings of the qualitative analysis conducted for the pre- and post-evaluation phases of the interviews, which propose a comparison of the participants' ideas throughout seven stimulus pairs. The analysis of the pre-evaluation phase (section 5.2.1) presents participants' design strategies when using only a video as a stimulus. The analysis of the post-evaluation phase (section 5.2.2) presents participants' design experiences when movement-based features were available. The results of the third part of the interviews (i.e., the overall experience, limitations faced, and the feature convenience in design practice) are further presented as part of the discussion section (section 5.3).

5.2.1 Design Strategies Using Only Video

The main themes that emerged from the participants' observations of human movement using only videos account for different design strategies: First, through the use of *existing interactive experiences* as design examples. Second, expressing the *detection needs* that would need to be resolved. Third, evaluating the *convenience of using user movement* to design interactive experiences. Finally, in certain cases, expressing original interaction design ideas. Table 5.1 shows an overview of the axial coding process, emergent categories, and themes.

Existing Interactive Experiences as Design Examples. Participants used examples of existing interactive experiences when asked for interaction design ideas in the context of a particular scene. This was a recurrent type of response at the beginning of the discussion and for all conditions, with the intent of using metaphors and references to familiar experiences in their own ideation process.

TABLE 5.1: Qualitative analysis of participants' design strategies using only video.

Axial Coding	Emerging Categories	Themes
Gesture Space Multi-modality Embodiment	Gesture control	Detection needs
User interest Grouping User characterization	User-space relation	
Movement patterns Collaboration	Group behavior	
References to prior work Design examples		Existing interactive experiences
Bodily descriptions Figurative descriptions Scene descriptions		Convenience of using user movement

> *"There is an interaction, from my initial perspective, that is very individual; as each one is on their own. So I thought like in these interactive control rooms that you see in the movies, I don't know, 'minority report' or something like that."* (P1)
>
> *"What comes to my mind are these experiences where you have a group of dancers in some kind of augmented performance, with projected images and mapping effects."* (P3)

In general, participants that used examples or references to prior works they have seen, respond with unresolved ideas and abstract design solutions. We relate this to the difficulty of identifying specific characteristics of the movement, in the absence of any visual feedback other than the video itself, to inspire design practice.

> *"It's like one of those 'infinite' interactive rooms, where I walk and something gets painted, and depending on how many people there are, then the painting changes."* (P5)
>
> *"Everything that has to do with using the gaze as a pointer. In fact, I forgot to mention: we used this in an experience at the Museum of Barcelona with the projection of a flashlight."* (P1)

Detection Needs. Participants were most concerned with using bodily motion as a natural user interface and expressed detection needs to be solved. As captured by some statements, their concerns pointed towards the *user-space relation*, *gesture control*, and *group behavior*.

> *"Well, you would have to identify the spaces where users stop. Then, if I can relate it to the direction they're looking at, I envision an immersive art experience."* (P6)
>
> *"I'm thinking about how long the user has to take to make the gesture and how to detect it. It is these types of parameters that at the end of the day would define the rhythm of the experience."* (P3)
>
> *"For me, the most interesting thing would be to embody the capabilities of a solo interaction, to see how much fun it is. How functional is what I am doing, just by moving*

> *that interface, moving those buttons? How involved am I in the interaction with the other person?"* (P2)

Convenience of using User Movement. When describing which aspects of users' motion characterize the scene, participants commonly found attributes closely related to the study conditions. Through this initial discussion, we validated that all scenes had valuable situations to further assess the prototype feature proposals. If these descriptions referred to the use of the body they were grouped as *bodily descriptions*:.

> *"A user went much faster, crouched down, had items to grab and to put from one side to the other. While the other was clearly looking around, recognizing the place, using very little of his body."* (P4)
>
> *"There are small gestures. I don't know, one user touches the other's shoulder, or some are looking at each other face to face. But I think the main feature in the video is the way users come together in a social movement."* (P3)

Statements describing overall behaviors or time and space circumstances were labeled *scene descriptions*:

> *"I note that they do not see each other, I do not see a correlation between their movements. I do see them occupying the same space, without there being a dissonance between what they are doing. However, I don't see a direct interaction between what they are doing there."* (P5)
>
> *"I think what matters most in that scene is where they stand and the interactions they are making in mid-air."* (P2)

While descriptions that used creative assumptions to interpret user movement were labeled as *figurative descriptions*:

> *"On the one hand, the first user enters, with an attitude of observing as if he were looking through a window. There isn't much more interaction than trying to see something from different angles. On the other hand, when the other users arrive, I imagine a situation where they try to find that thing that catches the attention of the first user."* (P1)
>
> *"It could also be an experience based on detecting if the user follows a path, let's say something like the circular movement of users. I imagine a control system that differentiates movements in large circles from smaller circles to run a choreographic show."* (P6)

5.2.2 Design Strategies Using Movement-Based Features

The review of the scenes with the assistance of the feature extraction and visualization prototype (see section 4.2) focused the interaction design strategies of the participants on the following themes: First, *feature interpretability*. Second, the *ease of use of features*. Third, the *proposal of new features* once they see what the prototype can offer. And finally, *imagining new forms of interaction* as a comparison of what they already expressed in the absence of any visual feedback other than the video itself. Table 5.2 summarizes the qualitative analysis process and how axial coding, categories, and themes emerged.

TABLE 5.2: Analysis of participants' design strategies when using the prototype.

Axial Coding	Emerging Categories	Themes
Motion interpretation Acceptance and endorsement Tool for analysis		Feature interpretability
User's sense of agency Natural user interface	From the user's perspective	Ease of use of features
Visualization impact Implementation issues	During development	
Feature parametrization Better extraction strategies	Algorithm refinement	Proposals for new features
Predict emotions Predict behaviors	Using features for prediction	
Simple derivations Time influence Relations with space	Higher-level features	
Interaction control User following	User journey	Imagining new forms of interaction
Creation and design Concept ideation	Promote the ideation process	
Prototype functionality		
Games Exhibitions	Applicability of the features	
Directions for moving Relation with time Motion description	Movement and time influence	

Feature Interpretability. Statements within this theme were further analyzed and clustered into the following sub-groups:

Motion interpretation consisted of descriptions of how the prototype feedback favored the practitioner's interpretation of motion and influenced their design ideas.

> *"Speaking of visualization, I have not only the size of the cluster, which in this case represents the number of users, but also the distance between those users to the cluster centroid, which can be very interesting to have different representations of the interaction that is happening."* (P4)
>
> *"One thing is that this feature allows me to see something different from what I saw in the corresponding video. The other is that the volume occupied by the user changes showing some very interesting patterns that I can use directly to control an interactive experience."* (P3)

Acceptance and endorsement were participants' descriptions of how they received the prototype functionality from an interpretability perspective.

> *"This is very powerful. Even more so with the possibility of choosing any landmark to detect similarities in the trajectory. I would choose hands, for example."* (P5)

> *"Interesting! The group movement in this visualization looks natural and fluid. I would use it not only to implement an interactive experience but to measure its effectiveness."* (P3)

Tool for analysis consisted of descriptions that referred to the prototype as a helpful resource to evaluate not only users' movement but the interaction experience over time.

> *"In this scene, it was 'cool' to see a heat map with a lot of the information that I had previously with lower-level features. It's all there, and now I can analyze where they were moving! It is certainly easier for me to see it because of the type of visualization."* (P3)
>
> *"For me, it is useful not only to control the experience as such but also to measure its long-term success. This is one of the most difficult things in the design of interactive experiences since in most cases the only feedback we have is through satisfaction interviews."* (P1)

Ease of Use of Features. Participants had differing opinions about which aspect of the use of the feature would make their design practice easier. On the one hand, some participants stated that it greatly depends on users understanding the movement they should make, which talks about considerations *from the user's perspective*. On the other hand, for some others, it depends on designers being able to map them as a control parameter, which represents considerations *during development*. The two following comments reflect previous postures, respectively:

> *"When you're designing experiences, it's important that there are certain things that are kind of magical and abstract, right? But also that they allow some understanding of how user interaction works in order to control the experience. Otherwise, people come in, and if they don't understand, they leave. So it also seems to me that it is, in a certain way, a matter of easy understanding for the user."* (P1)
>
> *"In clustering, I see that there are several data available that are, in a certain way, parameterizable. Not only for the user, as an interaction parameter, but as a robust control event in the implementation stage."* (P5)

Proposals for New Features. During the interviews, we noticed that as participants gained confidence in their understanding of the features, their comments on the evaluation of whether user movement was interesting for interaction design reflected the need of extracting information with a higher level of abstraction to meet the expectations of their own design ideas. Within this theme, three groups emerged:

Algorithm refinement gathered comments that evidenced that participants felt familiar enough with the extraction technique to envision their own ideas of using the algorithm to further their design ideas.

> *"I would use the user's distance to all other clusters in a game to influence what the other groups of people produce even if to a lesser extent."* (P4)
>
> *"Something that occurs to me, at a technical level, is to use several of the features that we have talked about as input to the machine learning algorithm to explore multidimensional clustering."* (P5)

Using features for prediction consisted of descriptions pointing towards the use of the extracted information to predict users' emotions or behaviors, although none of them directly mentioned how to achieve this computationally.

> *"Having access to the volume that each user is occupying allows me to predict their emotions, or what they are feeling. Even assign them a label of what type of user they are."* (P4)
>
> *"There is something interesting, and it is to understand that the feature is a specific information of the movement that opens the additional possibility of predicting the user's behavior."* (P5)

Higher-level features were responses that took a broader view and expanded the detection capabilities by using the given feature to propose a higher level of abstraction. In most conditions, the feature visualization helped participants to get an idea of how the temporal or spatial variability of the descriptor represented interesting detection events.

> *"It seems to me that it would be very interesting to relate the 'trajectory feature' over time. Detect things like a user who was still for a certain amount of time or how long the synchronized movement of users lasted."* (P1)
>
> *"Well, I have the 'feature' but its variability obviously seems interesting to me as another way of detecting something. For me, the ideal would be to describe the scene from information derived from the visualization of this feature."* (P6)

Imagining New Forms of Interaction. When asked what other interactions they could think of, participants seemed to have reasoned the feature as design material, and most were able to settle their thinking with concrete design ideas. The two most important categories from this theme grouping were: (1) descriptions of how the feature helps define, follow, or control the *user journey* throughout the experience.

> *"The interior design of commercial spaces has some rules, right? Then, depending on the user's trajectory, the intent of those rules can be reinforced with lighting, for example. Now I can detect where the user stops and, based on this event, understand which are the objects, or at least the category of objects, that attract the most attention."* (P6)
>
> *"It occurs to me that you can know two people from the same team, in what position they are one with respect to the other. And define whether or not they are fulfilling a condition, for example, you are 'offside' or you are not. You're ahead or you're not, those kinds of interactions. I think they are interesting."* (P2)

And (2) descriptions of how the prototype *promotes the ideation process:*.

> *"It can also lend itself to generative art explorations based on x number of rules. I would use as initial input, the trajectories of the hands as if they were pointers."* (P4)
>
> *"If the user's 'heading' is in this direction, then I light only this section, and the rest I make it dark. That way the user can use his body as a flashlight, to explore."* (P1)
>
> *"Imagine dividing the interaction space into two parts. You put a point of attention on each side, with a virtual button located in mid-air. You can identify when both users are pressing the button simultaneously, without having to consider a physical button as a sensor for such an event."* (P4)

To a lesser extent, the discussion of the participants focused on proposing interaction design ideas derived from the *Prototype functionality*, the *applicability of the features*, or after abstracting the *movement and time influence* through visualization.

> *"The information provided by the prototype can be used to develop an active play installation using gross motor skills. You could analyze group patterns and make a game out of them."* (P2)
>
> *"For example, I imagine a game where people have a screen in front of them and they have to move on a kind of platform, like a pinball. Then they have to become small and large to go through a maze. So users move around in space and have to talk about: 'become small', 'you're very heavy.' This is enough to get into a pretty fun interaction."* (P4)
>
> *"I immediately think of a game between movement and time, that is, if the user spends a lot of time moving a lot, then a whirlwind is generated, a little earthquake. So the user feels an earthquake in his body and wants it to start shaking. He needs a very long tail, right? The intensity would be directly correlated with the length of the tail. If he stays still, then the earthquake stops."* (P1)

5.3 Discussion

5.3.1 Hypothetical Scenarios for using Multi-view 3d Pose Estimation and Tracking

In this section, we explore hypothetical scenarios for applying multi-view 3d pose estimation and tracking in interaction design. Interaction designers were prompted to envision situations in which this technology could offer advantages for analyzing movement in interaction design contexts. The ensuing discussions uncovered various themes, including the impact of *group size,* the *timing* of technology utilization, and considerations related to the *sensing technology*.

Group Size. One key theme that emerged from our discussions with interaction designers is the significance of group size in determining the utility of multi-view 3d pose estimation and tracking. Participants unanimously agreed that the prototype would be most beneficial in scenarios involving a large number of users. This finding aligns with the recognition that, in settings with a considerable number of participants, manually analyzing and comprehending the extensive amounts of data generated by the system can be a daunting task. As P4 articulated, *"In massive use settings, where users are expected to pass quickly, cameras should be used. For example, in a museum, an event, or a concert."* Traditional technologies such as motion capture may excel in controlled environments, but they struggle to keep pace with the dynamics of large, unstructured audiences. Multi-view 3d pose estimation and tracking, through its clustering features, can serve as a valuable tool for designers by automatically identifying patterns and grouping users. This capability promotes the exploration of group behavior, allowing designers to identify key interaction opportunities and challenges rapidly. Such insights inform the development of more effective and engaging interactions for larger user groups. Furthermore, the visualization of clustering features aids in comprehending the flow of people, as noted by P6: *"Without visualization, it is very difficult for me to really know how interaction groups* *are formed."* Visualizations of hotspots reveal areas of high user interaction, particularly crucial in scenarios with a large number of participants. Designers can leverage this information to optimize system design, enhancing the overall user experience.

Timing. The timing of employing multi-view 3d pose estimation and tracking appears as another vital consideration. Participants in our study acknowledged that this technology could be an invaluable resource at various stages of interaction design. For instance, some participants suggested that it could aid in the early ideation phase by providing insights into how users naturally move (P1 and P3). This information can fuel the generation of innovative interaction ideas. However, for real-time applications, participants emphasized the need for improved synchronization and faster processing, highlighting the importance of refining the technology's performance (P6). Additionally, the prototype was seen as a potent analysis tool, enabling the long-term offline examination of movement data (P2, P4, and P5). This extended temporal perspective can be instrumental in maintaining or redesigning interactive experiences based on comprehensive movement data analysis.

Sensing Technology. Interviewees discussed the advantages and disadvantages of other sensing strategies concerning accuracy, robustness, latency, and complexity, comparing them with the multiple-view approach presented in this study. A recurring theme in our discussions was the preference for implementation simplicity over the pursuit of extreme accuracy in interactive system development. Participants appreciated the ease of using consumer cameras and the adaptability of spatial configurations. These technologies were perceived as user-friendly and accessible, facilitating a seamless integration into various design scenarios. This aligns with the ideas expressed by P2: *"It is an advantage to have cameras. I find the fact that large spaces can be covered very interesting."* In contrast, more accurate sensing setups, such as motion capture systems and inertial measurement units, were acknowledged for their potential to deliver precise data. However, their adoption was hindered by the challenges they posed. The key factor at hand is striking a balance between precision and practicality. As P2 emphasized, *"Obviously, its use comes with a sacrifice issue in the precision of the movement."* This acknowledgment underscores the pragmatic approach taken by interaction designers in our study. None of the interaction concepts discussed required fine-grained movement tracking or sub-millimeter accuracy. Instead, the emphasis was on capturing the broader patterns and dynamics of user movement, which aligns well with the capabilities of multi-view 3d pose estimation and tracking.

5.3.2 Feature Convenience in Interaction Design Practice

Interaction Designers had different opinions on which of the features presented would be most suitable for their interaction design practice. However, all participants agreed to select features they thought would make the interaction experience functional and easy to replicate. This implies preferring features that would contribute to creating an interaction that is practical, usable, and efficient.

We observed a split in participants' attitudes about what determines functionality, with the difference coming down to who would directly benefit from the detection strategy. On the one hand, four of the six interviewees stated that they would choose a feature based on whether end users understand the movement they should make, thus providing users with a greater sense of agency. P1 argued, *"I would choose* *a feature that helps users understand what they need to do without too much instruction so that the interaction feels natural and intuitive."* Additionally, P3 mentioned, *"For me, a functional interaction is one that is easy to learn, easy to use, and doesn't require a lot of effort from the user. So, I would select features that simplify the interaction and make it more straightforward."* According to the practitioners, features such as Bounding Box, Heading, and Instantaneous Clustering were viewed as dimensionality reduction strategies that could simplify the design of an interaction strategy, particularly when designers needed to minimize the time users spent learning how to use the system. On the other hand, the two other interviewees felt that they could potentially benefit from movement-based features if they could map them reliably as a control parameter during implementation. P6 expressed, *"I think the interaction needs to be reliable and consistent so that users can trust that it will work the same way every time. Easy replication means that I can easily implement the interaction in different contexts without a lot of customization."* At face value, designers expressed the features' attributes that better suit their practice: coarse actions characterization, user motion anticipation, and association of detection events with natural and intuitive movements.

These findings hold broader implications for the research community, shedding light on the decision-making process of interaction designers. It underscores the need for researchers to develop a nuanced approach to feature design, recognizing that designers weigh considerations of end-user experience and practical implementation differently based on their specific project goals. Consequently, tools and guidelines should be flexible and adaptable, accommodating designers' diverse preferences and priorities. Moreover, our study accentuates the critical role of perceived agency by end-users as a driving factor in practitioners' decisions when incorporating movement-based features into their interaction design practice. This emphasis on empowering users reflects a broader trend in the field of interaction design, where interactions are increasingly designed to prioritize user engagement, satisfaction, and intuitive control. In an evolving digital landscape, users have heightened expectations, seeking interactions that not only fulfill functional requirements but also empower them to effortlessly achieve their goals.

However, it is crucial to distinguish our findings from existing research that has delved into the use of movement qualities features. These studies have often been conducted within expert domains, involving individuals with extensive training in the nuances of movement performance and the subtleties of movement qualities. In these specialized contexts, the focus lies on optimizing the movement representations, serving experts with a deep understanding of movement nuances. In contrast, our study addresses the broader field of interaction design, where the end-users are often non-experts with varying degrees of familiarity with the digital interface. Our findings underscore that interaction designers prioritize perceived agency by these end-users, which may not necessarily align with research objectives conducted in expert domains. Interaction designers aim to create intuitive, user-centric experiences that empower individuals of diverse backgrounds, potentially with minimal training or expertise in movement qualities. Therefore, our results emphasize the practical considerations and user-centric approach that guide feature selection in interaction design, distinct from the specialized focus of certain related research areas. This distinction highlights the relevance and applicability of our findings in the broader context of designing interactive systems for a wide range of users.

5.3.3 Feature Visualization in Designing Movement-Based Interactions

Our research bridges artistic exploration and analytical needs through feature visualization in movement-based interactions. In contrast to conventional visualization methods, which often fall short of capturing nuanced movement dynamics, our approach provides tailored visualizations for interaction designers. While prior work [88], [143] has predominantly focused on enhancing expressive and enjoyable aspects of movement, our research takes a distinct analytical stance. We empower a diverse audience, including non-experts, by making complex movement data comprehensible and expanding the reach of movement-based interaction design to various applications. By acting as a bridge, our work serves as a valuable tool for data-driven decision-making across diverse domains. It complements the existing landscape of movement visualization techniques by addressing the specific needs of interaction designers. Furthermore, our approach extends beyond contemporary dance and expressive gesture recognition [144], [145], finding applications in user experience design, smart learning environments, public installations, and more.

Practitioners reacted positively to the designer-interpretable features and how their visualization could improve their interaction design practice. The main benefit practitioners perceived was that the prototype allowed them to explore interaction ideas at first glance, without complex interpretation barriers or prior training before using users' movement during ideation. Visual cues helped them to contextualize users' individual and group intentions in space, laying the groundwork for (1) enhancing the design process and (2) understanding user experience.

5.3.3.1 Enhancing the Design Process

Our study reveals insights into how the integration of feature visualization enhances the design process in the context of movement-based interactions. Beyond the initial positive reactions from practitioners, it's essential to understand the underlying mechanisms that drive this enhancement.

Accessibility and Ideation. The intuitiveness of our prototype and the immediate accessibility it provides to practitioners offer crucial insights into the removal of barriers in the early stages of interaction design. This newfound ease of entry into the design process fosters creativity. It encourages practitioners to explore unconventional ideas and take risks, knowing they can quickly translate ideas into visualizations. This is particularly valuable in the ideation phase, where the freedom to experiment without inhibition can lead to innovative design concepts.

Validation and Iteration. The ability to compare design ideas with initial perceptions through feature visualization is a powerful validation mechanism. Designers can see how closely their conceptualizations align with the actual movements and interactions, which not only aids in refining ideas but also facilitates iterative design. By continuously comparing design iterations, practitioners can identify areas for improvement and make informed adjustments. P3 argued, *"It's just that the features make it much easier. For example, clustering and trajectories are good for understanding how groups are formed. One now compares the features and the videos, and it is not easy to realize all the information in the movement that can be useful."* Thus, our approach provides a feedback loop that promotes iterative and user-centered design practices.

Facilitating Cross-disciplinary Ideation. One noteworthy insight is the capacity of our prototype to transcend disciplinary boundaries. As stated by P1, *"The prototype is very useful for me to ground the idea and take it to other fields of ideation."* It suggests that feature visualization doesn't just benefit interaction designers but has the potential to support cross-disciplinary collaboration. It acts as a common language that allows professionals from diverse backgrounds to communicate and iterate on ideas effectively.

5.3.3.2 Understanding User Experience

Our study also provides valuable insights into how feature visualization aids in understanding user experience in movement-based interactions, going beyond immediate benefits to provide deeper understanding.

User-centric Design. The ability of designers to imagine and simulate user interactions through visual cues represents a shift toward more user-centric design practices. By offering a visual representation of user movements and group dynamics, our approach encourages designers to think from the user's perspective. This empathetic viewpoint enables designers to anticipate user needs, pain points, and preferences, resulting in designs that are more aligned with user expectations.

Concrete Decision-making. The insights gained from visualizing features, such as distance to centroids, angles between users, and cluster shapes, provide designers with tangible and actionable information. This information helps them make informed design decisions. For instance, as mentioned by P5, *"By observing a trajectory, I can be very sure of where the user has passed, and I can make redesign decisions with certainty based on an analysis over time."* This concreteness empowers designers to create interactions not based solely on intuition but grounded in data-driven insights.

Overcoming Computational Complexity. Practitioners reported that exposure to visual cues instead of numerical data favored their design practice. Showing a few spatial and temporal relationships derived from the algorithm's output helped participants consolidate a graphical understanding of the feature. An essential aspect highlighted by P3's comment is how feature visualization simplifi*"What is very difficult to identify without visualization are, for example, the creative possibilities of using clustering and those features that are a bit more computational and therefore harder to interpret."* Movement-based interaction data often involves intricate calculations, making it challenging for designers without a technical background to grasp the underlying concepts. Visualization acts as a bridge, converting abstract numerical data into easily interpretable visual representations, promoting understanding and teamwork among designers, data scientists, and engineers.

5.4 Summary

Chapter 5 offers an in-depth exploration of an innovative approach centered on visualizing user motion descriptors within the context of interaction design. The prototype introduced in chapter 4.2 presents a transformative tool that empowers interaction designers to harness the potential of movement data effectively. By automating feature extraction, the prototype facilitates the exploration of diverse techniques, aiding designers in identifying optimal approaches aligned with their design objectives. A key highlight lies in the prototype's visualization capability, which uncovers intricate patterns and relationships hidden within the raw data, enriching designers' understanding.

Further enriching the design process, the chapter underscores the critical role of visualization in tailoring interactions to user behaviors and movement patterns. Our study confirms that the approach enhances designers' use of human motion data and fosters creativity in design. The prototype's potential extends to improving the comprehension of user experiences, offering a clear visual representation of extracted features. Moreover, this method expands the scope of feature extraction methods, which is especially useful when dealing with big and intricate datasets that present difficulties for manual analysis. Ultimately, the chapter accentuates how visualization can enhance interaction design by infusing it with profound insights from motion data.

Chapter 6

Conclusions and Future Work

In this chapter, we reflect on the current research in light of the thesis objectives, the specific purposes of three publications, and the problem outlined in the first chapter of the thesis. Similarly, we highlight the limitations of the study and suggest some avenues for future research.

6.1 Contributions of the Thesis

This thesis makes significant contributions to the field of movement-based interaction by addressing the challenge of designing user-centered, motion-based interactions. Through a series of studies, it explores the perceptions of interaction designers towards using motion-based full-body features, proposes a set of *designer-interpretable* features that enhance the designer's ability to use human motion in interactive experiences, and evaluates the effects of these features on the concept ideation and design of motion-based interactions using a user-centered approach.

The first contribution of the thesis is the exploration of the perceptions of interaction designers towards using motion-based full-body features. This contribution is particularly important because interaction designers are the ones responsible for designing interactive systems that take advantage of human motion. However, the use of motion-based full-body features is an emerging area of research within interaction design, and little is known about how designers perceive and approach this kind of technology. The second major contribution of the thesis is the development of a set of *designer-interpretable* features that can be used to enhance the designer's ability to use human motion in interactive experiences. These features were developed based on the results of a focus group study that aimed to understand the perceptions of interaction designers towards using motion-based full-body features and how these features link sensor-based interaction design with feature extraction technology. The third contribution of the thesis is an evaluation of the effects of the proposed features on the concept ideation and design of motion-based interactions following a user-centered strategy. The validity of the approach is established through a qualitative study conducted with interaction designers.

The specific objectives of the thesis plan were concretized as the research progressed, taking the form of specific purposes in the three mentioned papers: A.1 investigate the perceptions of interaction designers towards the use of motion-based full-body features in their design work, A.2 explore the impact of user clustering visualization on the design of motion-based interactive systems, and A.3 propose and evaluate a user-centered approach for using movement-based features in interaction design.

6.2 Conclusions

Chapter 3 offers an integral vision of the challenges and opportunities that practitioners encounter when processing sensor data to extract motion-based features for interactive experience design. The study with interaction designers allows us to argue that there are various designer profiles, and that some may not have a solid technical background. This divergence can limit the full potential of feature extraction technologies. This thesis has thoroughly explored this critical distinction and its significance. By carefully outlining the interaction designer's profile, this research has shed light on why different designers engage with complex algorithms and technologies to varying degrees. As a result, this chapter not only enriches the understanding of the intricate relationship between designers and technology but also provides a comprehensive perspective on why some designers may not be as involved with these innovations as others.

Additionally, Chapter 3 highlights that while feature extraction is considered challenging and time-consuming, a computational tool designed to interpret the movement of multiple users from diverse perspectives could enable novel interaction strategies to emerge at a reduced technical cost. Furthermore, it emphasizes that motion-based features should be grounded in conceptual frameworks known for offering advice on designing sensor-based interactions. The chapter adheres to the idea that non-technical interaction designers must understand the possibilities opened up by technology and acknowledge the detection system's nature to complete an effective design-development cycle. The study in this chapter shows that acknowledging the detection system's nature is crucial for interaction designers, even if they lack the technical knowledge to extract features. Overall, this chapter widens the reach of feature extraction technology for motion-based interactive experiences by demonstrating its potential as a useful tool for better understanding user behavior. The study also validates the concept of a computational tool designed as a *designer-interpretable* motion-based feature extractor, which offers a comprehensible representation of the interaction design possibilities that sensing technology offers.

Chapter 5 presents an approach to visualizing user motion descriptors, providing interaction designers with a valuable tool to effectively utilize movement data in the design process. By automating the process of feature extraction, our computational prototype allows designers to quickly iterate through different techniques and identify the most relevant ones to their design goals. The visualization allowed designers to overcome the complexity of the extraction algorithms and explore the creative possibilities of using motion-based features. Additionally, the prototype makes it easy for designers to identify patterns and relationships that are not immediately apparent from raw data. By identifying common behaviors and movement patterns of users, designers were able to tailor their interactions accordingly, resulting in more engaging and interactive experiences for end users. The exposure to visual cues, rather than numerical data, was deemed crucial in their design practice as it helped them to understand user behavior in a more intuitive and graphical way. Practitioners reported that it was very difficult to abstract the meaning of the data without the prototype, and by observing the user trajectory, they were able to make redesign decisions with certainty based on an analysis over time. The results of the qualitative study have shown that this approach is effective in improving the ability of interaction designers to use human motion data in the design process, leading to more creative and innovative design ideas, as well as more engaging and interactive experiences for end-users. Furthermore, the prototype allows designers to better understand the user experience by providing a clear and intuitive representation of the features extracted from human motion, enabling designers to better understand how users are interacting with technology. One of the key benefits of this approach is that it widens the reach of feature extraction techniques for interaction designers, particularly when working with large and complex datasets, where manual analysis can be time-consuming and error-prone. Additionally, by providing a visual representation of the algorithm's output, designers can easily identify the abilities and limitations of the detection system, which can inform the design process in a way that can lead to novel interaction ideas.

6.3 Limitations

This research is focused on a specific domain, namely the design of movement-based interactions. While the findings and approaches presented in the thesis are relevant to this domain, their applicability to other domains may be limited. The results of the studies may not be generalizable to other types of interactions or design domains. Moreover, the studies primarily focus on the design phase of the interaction design process. While this is an important phase, it is only one part of the overall design process. The proposed approaches and prototypes may need to be integrated into other phases of the design process, such as evaluation and testing, to fully realize their potential benefits.

The studies conducted in the thesis were limited to a relatively small sample size of interaction designers. While the results are promising, further research with a larger and more diverse sample may be necessary to confirm the findings and generalize them to a broader population. Additionally, the prototype developed in the studies have limitations in terms of their functionality and scalability. While the prototype was effective in demonstrating the feasibility and potential benefits of the proposed approaches, it may require further development and refinement to be implemented on a larger scale or in real-world design projects.

The study in Chapter 5 focuses on specific aspects of movement-based interactions, such as user clustering and trajectory similarity, to describe group behavior (see section 4.2). While these are important areas of research, there may be other important aspects of movement-based interactions that have not been explored in this research. Further research may be necessary to identify and investigate these aspects.

Practitioners should be aware that motion-based feature extraction is a complex process that requires an understanding of the domain, the user, and the technology. It should be approached with a critical perspective, considering the context, the goals, and the limitations of the project. Also, it's important to note that user behavior visualization should be used in conjunction with other evaluation methods, such as usability testing, to provide a more comprehensive understanding of the system's performance.

6.4 Future Work

In future work, several aspects related to the contributions of the thesis can be further explored. One important direction is to investigate the impact of user clustering visualization on the design of interactive experiences in more depth. To this end, user studies can be conducted to evaluate the effectiveness of the system in different scenarios and domains. Moreover, the integration of other machine-learning techniques can be explored to improve the accuracy and robustness of the user clustering algorithm. As a new line of research, one can investigate the potential of using our approach in other fields such as crowd analysis, sports analysis, and surveillance. In particular, our approach has the potential to provide insights into movement patterns and behaviors of people in different contexts, which can be useful for a wide range of applications.

Another area of future work is to explore the potential of using the computational prototype's approach in combination with other movement-based interaction design techniques such as gesture recognition. This would enable the creation of more complex and natural interactions for users, and further improve the accuracy and robustness of the system. Additionally, the use of virtual and augmented reality technologies can be investigated to enhance the visualization of group behavior descriptors and provide a more immersive design experience for interaction designers. This can be particularly useful for designing complex interactive experiences in domains such as entertainment and education.

Finally, the computational prototype's approach to visualizing user clustering descriptors can be employed in a variety of different contexts. For example, it can be used to design interactive installations, games, and other types of interactive experiences. It can also be used in fields such as healthcare, where it can be used to track and analyze the movement of patients in order to better understand and treat conditions such as Parkinson's disease and other movement disorders. Thus, the potential applications of our approach are diverse and promising, and further research is needed to explore its full potential.

Appendix A

List of Publications

A.1 Interaction designers' perceptions of using motion-based full-body features

Escamilla, A., Melenchón, J., Monzo, C., Morán, J.A., 2021. Interaction designers' perceptions of using motion-based full-body features. International Journal of Human-Computer Studies 155, 102697.

DOI: 10.1016/j.ijhcs.2021.102697.
JCR Impact Factor: 4.866 (2021), 1st quartile.
Category: COMPUTER SCIENCE, CYBERNETICS - SCIE.

Abstract

Movement-based full-body interactions are increasingly being used in the design of interactive spaces, computer-mediated environments, and virtual user experiences due to the development and availability of diverse sensing technologies. In this context, the role of interaction designers is to find systematic and predictable relationships between bodily actions and the corresponding responses from technology. Sensor-based interaction design relies on sensor data analysis and higher-level feature extraction to improve detection capabilities. However, understanding human movement to inform the design of motion-based interactions is not straightforward if the detection capabilities of interaction technologies are unknown. We aim at understanding the problems and opportunities that practitioners—regardless of their technical background—perceive in using different motion-based full-body features. To achieve this, we conducted four separate focus groups with experienced practitioners, with and without technical backgrounds. We used a framework for the analysis of focus group data in information systems research to identify content areas and draw conclusions. Our findings suggest that most interaction designers, regardless of their technical background, consider motion-based feature extraction to be challenging and time-consuming. However, participants acknowledge they might use *designer-interpretable* features as a potential tool to foster user behavior exploration. Understanding how practitioners link sensor-based interaction design with feature extraction technology is relevant to design computational tools and reduce the technical effort required from designers to characterize the user's movement.

A.2 User Clustering Visualization and Its Impact on Motion-based Interaction Design

Escamilla, A., Melenchón, J., Monzo, C., Morán, J.A., 2023. User Clustering Visualization and Its Impact on Motion-based Interaction Design. In: Kurosu, M., Hashizume, A. (eds). HCII 2023. Lecture Notes in Computer Science, vol 14011. pp. 47–63. Springer, Cham.

DOI: 10.1007/978-3-031-35596-7_4
Conference: 25th International Conference on Human-Computer Interaction, HCII 2023.

Abstract

Movement-based interaction design relies on sensor data analysis and higher-level feature extraction to represent human movement. However, challenges to effectively using movement data include building computational tools that allow exploring feature extraction technology as design material, and the need for visual representations that help designers better understand the contents of movement. This paper presents an approach for visualizing user clustering descriptors to enhance the practitioners' ability to use human motion in interaction design. Following a user-centered strategy, we first identified perceptions of, and barriers to, using motion-based features in a group of interaction designers. Then, a multiple-view multiple-people tracking system was implemented as a detection strategy that leverages current models for 3d pose estimation. Finally, we developed a computational prototype that performs instantaneous and short-term clustering of users in space and presents simple descriptors of the algorithm's output visually. Our approach was validated through a qualitative study with interaction designers. Semi-structured interviews were used to evaluate design strategies with and without the assistance of the computational prototype and to investigate the impact of user clustering visualization on the design of interactive experiences. From practitioners' opinions, we conclude that feature visualization allowed designers to identify detection capabilities that enriched the ideation process and relate multiple dimensions of group behavior that lead to novel interaction ideas.

A.3 Exploring a user-centered approach to using movement-based features for interaction design purposes

Escamilla, A., Melenchón, J., Monzo, C., Morán, J.A., Carrascal, J.P., 2023. Exploring a user-centered approach to using movement-based features for interaction design purposes. (*Under review, ACM Transactions on Computer-Human Interaction*)

JCR Impact Factor: 3.7 (2022), 2nd quartile.
Category: COMPUTER SCIENCE, CYBERNETICS - SCIE.

Abstract

In this paper, we present an approach that considers the perspectives and attitudes of interaction designers toward incorporating motion-based features in their designs. We first develop a guiding framework that bridges the gap between low-level motion-based attributes and a designer-oriented representation. Secondly, we propose a set of *designer-interpretable* descriptors to enhance their ability to use human motion in interaction design. Finally, we use a computational prototype to visually present the features and help designers better understand movement content. The investigation into the effects of *designer-interpretable* features on the concept ideation and design of motion-based interactions followed a user-centered approach, and its validity was established through a qualitative study conducted with interaction designers. Semi-structured interviews were designed to play the role of a pre- and post-evaluation of design strategies when using the computational prototype and observe the impact of feature visualization on concept ideation. The study determined that interaction designers leveraged features to identify detection capabilities and enrich the ideation process. Moreover, feature visualization provided further insights into the characteristics of movement, which helped practitioners understand the interaction opportunities that come with it. The unanimous assessment of the proposed features as an effective tool for the analysis of interactive group experiences supports our approach and visualization strategy.

Appendix B

Focus Group Guide

B.1 Main and supplementary questions in each stage.

TABLE B.1: Main and supplementary questions in each stage.

Stage	Main Question	Supplementary Questions
Stage 1	How do you explore the different behaviors of users to develop an interactive experience?	What interaction design methodologies do you use most frequently?
Stage 2	How do you relate interaction design to the possibilities offered by detection technologies?	How do you translate detection systems outputs into a feature representation especially suitable for use in a creative context?
Stage 3	When using motion-based detection technologies, do you think about interpreting the sensor data?	What interpretation problems from sensor data have you usually found?
Stage 4	Which design needs should a set of features solve when creating motion-based multiple-users interactions?	Which design needs should a set of features solve in the context of large-volume spaces and full-body interaction? What features (information extracted from multiple-users movement) do you consider to be close to typical concepts of full-body interaction design in large-volume spaces? What features, drawn from multiple-users movement, would lead interaction designers to explore the body as a communication/interaction strategy?

Appendix C

Concepts in Focus Group Data Analysis

TABLE C.1: Frequently Used Concepts in Focus Group Data Analysis [111].

Concept	Definition
Content area	A content area is part of the text such as paragraphs or sentences each of which is about a similar concept, issue, etc., and are directly related to each other. In the literature, a content area has also been called a domain, rough structure, or cluster.
Manifest content	Parts (sentences, paragraphs, etc.) of the transcript and observation field notes with clear meaning (there is no need for a high level of interpretation) and meaning that multiple analysts agree with.
Latent content	Parts of the transcript and observation field notes that need a higher level of interpretation and require more discussion among research team to understand and agree on what the text talks about.
Meaning unit	Graneheim and Lundman (2004, p. 106) define meaning unit as “words, sentences or paragraphs containing aspects related to each other through their content and context.” This concept also has been called “idea unit,” “content unit,” “coding unit,” “textual unit,” and even “theme.”
Condensation	Shortening a text without changing the quality of its concept. Condensation also has been called reduction and distillation.
Abstraction	The process of grouping together the condensed text on varying levels such as codes and concluding subcategories, categories, and themes. Abstraction also has been called aggregation.
Code	A label/name, a colour, or a number assigned to a condensed meaning unit. In the analysis framework by code, we mean the use of labels in the analysis process.
Category	A group of similar codes, and may consist of a number of subcategories. ‘Category’ expresses the manifest content of the transcript and answers the question ‘What?’ As categories are exhaustive and mutually exclusive, no data can fit into more than one category and no data must be excluded due to lack of an appropriate category.
Theme	“A thread of an underlying meaning through, condensed meaning units, codes or categories, on an interpretative level” (i.e., the expression of the latent content). A theme may include subthemes and answers to the question ‘How?’. As the measure not necessarily mutually exclusive, one or more condensed meaning units and even codes and categories may fit into more than one theme.

Appendix D

Interview Protocol

1. The first part of the stimuli, a multi-view video with no additional information, was presented. The participant assessed whether people's movement is of interest for interaction design using just the video and discussed design ideas for spatial-interactive experiences in the context of the scene shown. The following specific questions were made:
   - Which aspects of people's motion characterize the scene?
   - How would you imagine an interactive experience in the context of a scene like the one shown?
2. The second part of the stimuli pair is presented. The participant reviewed the same scene, having visual assistance from the feature extraction and visualization prototype, and answered the following questions:
   - How do you interpret the people's motion from the extracted feature?
   - What other interactions can you think of with the information you have now available?
   - How would you use the extracted information available to design a spatial-interactive experience?
3. Steps 1 and 2 were repeated for all seven stimulus pairs, following up on topics of interest that arose naturally.
4. As a closure, after all stimuli had been discussed, the participant was asked about aspects that influenced their overall experience, their thoughts on the difficulties of assessing people's movement without proper visualization, and the limitations faced. The following specific questions were made:
   - Before adding the feature visualization, how easy was it to assess people's movement from the video? Comment on this, please.
   - Among all the features presented, which ones seem convenient to your design practice and why?
   - Can you formulate hypothetical scenarios where you would see benefits from using 3d pose and tracking to analyze people's motion for interaction design purposes?

Appendix E

Concepts in Grounded Theory Analysis

TABLE E.1: Frequently Used Concepts in Grounded Theory Analysis [116].

Concept	Definition
Open coding	Open coding is the initial stage of qualitative data analysis in Grounded Theory. It involves breaking down the raw data (interview transcripts, field notes, etc.) into discrete segments or "chunks" and assigning descriptive labels or codes to these chunks. Open coding aims to identify concepts, patterns, and ideas within the data without any preconceived categories. It's about creating meaningful labels that capture the essence of the data.
Axial coding	Axial coding comes after open coding and involves organizing and connecting the codes identified in the open coding phase. It's about exploring relationships between codes and developing a more structured understanding of the data. Axial coding often involves creating diagrams or visual representations to show how codes relate to each other, leading to the emergence of themes and categories.
Code	A code is a label or tag that is assigned to a specific segment of data (such as a sentence or a paragraph) to capture its meaning or significance. Codes can be descriptive (representing the content of the data) or conceptual (representing the underlying ideas or concepts). Codes are the building blocks of analysis and are used to identify patterns and trends within the data.
Emergent category	Emergent categories often result from the coding process. Researchers initially code the data by breaking it down into smaller segments or codes. As they analyze and compare these codes, patterns and similarities emerge, leading to the development of categories.
Theme	A theme is a central idea, pattern, or concept that emerges from the data analysis. It represents a recurring topic, concept, or phenomenon that is relevant to the research questions. Themes are identified through the process of open and axial coding and help researchers make sense of the data by grouping similar categories together under overarching concepts.
Label	A label refers to a short and concise descriptor or term that is assigned to a code, category, or theme to succinctly capture its essence or meaning. Labels help researchers quickly identify and reference specific concepts within the data. Labels should be chosen thoughtfully to accurately represent the content or concept being referred to.
Theory	In Grounded Theory, a theory is not necessarily a formal theoretical construct but rather an explanation or conceptual framework that emerges from the data analysis process. The theory is grounded in the data and represents an understanding of the phenomena being studied.

Appendix F

Script for a group of actors

F.1 General Instructions

- **Scene Overview:** We will be performing a series of seven social interaction scenes, each specifically designed for a particular motion-based feature. These scenes will serve as stimuli for our study.
- **Scene Depiction:** In these scenes, you will portray four actors engaged in interactions set within motion-based environments. The recordings from these interactions will result in video clips, each lasting two minutes.
- **Behavioral Freedom:** There are no strict behavioral constraints imposed on you, except for a contextual scene description. This will allow for unscripted and authentic interactions, avoiding any artificial movements.
- **Scene Descriptions:** You will receive a strategic scene description for each scene, which will guide your performance. This description is crucial for aligning your actions with one of the seven proposed motion-based features.
- **Alignment Importance:** It's essential that your performances seamlessly match the chosen motion-based feature. This alignment will facilitate the integration of your actions with our computational prototype and help us extract valuable insights from your natural interactions.
- **Individual Movement:** Three out of the seven scenes will focus on exploring the intricate characterization of individual movements. This will highlight the nuanced essence of each actor's gestures.
- **Collective Behaviors:** The other four scenes will delve into collective behaviors, emphasizing spatial clustering and trajectory patterns within the interactions.
- **Interaction Dynamics:** Keep in mind that, within the context of motion-based interaction, the scene descriptions provide a foundation for your actions while allowing you to naturally respond to the movements and interpretations of your fellow performers.

F.2 Scene Descriptions

TABLE F.1: Scenes Descriptions and their Corresponding Feature.

Scene	Feature	Description
1	Bounding box	Movements of individuals with slow, expansive volume changes. Sometimes, they do this interchangeably or randomly, and at another moment, they do it in unison. They can simulate movements akin to freeform dance or in front of a responsive screen. There should be transitions to moments of minimal limb movement, maintaining constant the volume they occupy.
2	Trajectory	Movements of people walking randomly around the place. Crosses between them, some person stops. Some moments people try to move to generate a pattern in their trajectories.
3	Heading	A person enters the place and in the center begins to turn his body following a virtual object. At some point she stands still staring at the object and suddenly turns around and looks at the still object again. The same action is repeated as if something caught the attention behind the person.
4	Clustering	Four people meet at a central point, joining one by one, but they don't come together much. One person moves away from the group to another point and time later another individual joins this person. The two groups generate dynamics of social interaction, maintaining their average distance. From one of the groups, a person leaves the place and the person who is left alone looks for the other group to join them, after walking alone around the place.
5	Hotspots	One person walks up to a virtual wall, three others move around and are drawn to something that the first person appears to be looking at on that wall. One of them moves along this wall, in exploration mode and the other three add up. The same thing happens again around another point in space some time later.
6	Trajectory Similarity	Three people walk randomly and decide at some point to take a break looking at the ground. From different places they make the same movement following some virtual object on the floor that tells them where to move. It can be moving in a circle, around, or in a straight line with the same direction. At some point they move randomly until they meet in a line. They move back and forth.
7	Movement patterns correlations	Four people in front of a large screen make displacement movements and return to their starting position. Sometimes they take turns and do the same move with two or three of them, sometimes all four. The scene can be repeated based on the same concept, only now they do movements with the right hand instead of displacements.

References

[1] T. T. Hewett, R. Baecker, S. Card, *et al.*, *ACM SIGCHI curricula for human-computer interaction*. ACM, 1992, ISBN: 0-89791-474-0.

[2] E. Churchill, A. Bowser, and J. Preece, "Teaching and learning human-computer interaction", *interactions*, vol. 20, no. 2, p. 44, 2013.

[3] P. Dourish, *Where the action is*. MIT press Cambridge, 2001.

[4] G. Glonek and M. Pietruszka, "Natural user interfaces (NUI)", *Journal of Applied Computer Science*, vol. 20, no. 2, pp. 27–45, 2012.

[5] D. England, "Whole body interaction: An introduction", in *Whole Body Interaction*, Springer, 2011, pp. 1–5.

[6] D. Casillas-Perez, J. Macias-Guarasa, M. Marron-Romera, D. Fuentes-Jimenez, and A. Fernandez-Rincon, "Full body gesture recognition for human-machine interaction in intelligent spaces", in *International Conference on Bioinformatics and Biomedical Engineering*, Springer, 2016, pp. 664–676.

[7] A. B. Godbehere and K. Goldberg, "Algorithms for visual tracking of visitors under variable-lighting conditions for a responsive audio art installation", in *Controls and Art*, Springer, 2014, pp. 181–204.

[8] W. R. Pensyl, D. K. Jernigan, T. C. T. Qui, H. P. Fang, and L. S. Ping, "Large area robust hybrid tracking with life-size avatar in mixed reality environment: For cultural and historical installation", in *Proceedings of The 7th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and Its Applications in Industry*, ACM, 2008, p. 9, ISBN: 1-60558-335-9.

[9] P. Caserman, A. Garcia-Agundez, and S. Göbel, "A Survey of Full-Body Motion Reconstruction in Immersive Virtual Reality Applications", *IEEE Transactions on Visualization and Computer Graphics*, vol. 26, no. 10, pp. 3089–3108, Oct. 2020, ISSN: 1941-0506. DOI: 10.1109/TVCG.2019.2912607.

[10] G. Cosentino and M. Giannakos, "Multisensory interaction and analytics to enhance smart learning environments: A systematic literature review", *IEEE Transactions on Learning Technologies*, vol. 16, no. 3, pp. 414–430, 2023. DOI: 10.1109/TLT.2023.3243210.

[11] A. Chapman and N. Hume, "Random International", in *Coding and Representation from the Nineteenth Century to the Present*, Num Pages: 12, Routledge, 2021, ISBN: 978-1-00-316915-4.

[12] L. Loke, A. T. Larssen, T. Robertson, and J. Edwards, "Understanding movement for interaction design: Frameworks and approaches", *Personal and Ubiquitous Computing*, vol. 11, no. 8, pp. 691–701, 2007. DOI: 10.1007/s00779-006-0132-1.

[13] C. Bogdan, V. Tsaknaki, C. Windlin, *et al.*, “Programming for moving bodies”, in *Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society*, ser. NordiCHI ’20, Tallinn, Estonia: Association for Computing Machinery, 2020, ISBN: 9781450375795. DOI: 10.1145/3419249.3420069.

[14] H. Sharp, J. Preece, and Y. Rogers, *Interaction Design: Beyond Human-Computer Interaction*, 5th Edition. Wiley, 2019, ISBN: 978-1-119-54725-9.

[15] L. J. Bannon, “A human-centred perspective on interaction design”, in *Future Interaction Design*, A. Pirhonen, P. Saariluoma, H. Isomäki, and C. Roast, Eds. London: Springer London, 2005, pp. 31–51, ISBN: 978-1-84628-089-4. DOI: 10.1007/1-84628-089-3_3.

[16] V. Bellotti, M. Back, W. K. Edwards, R. E. Grinter, A. Henderson, and C. Lopes, “Making sense of sensing systems: Five questions for designers and researchers”, in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI ’02, New York, NY, USA: Association for Computing Machinery, Apr. 2002, pp. 415–422, ISBN: 978-1-58113-453-7. DOI: 10.1145/503376.503450.

[17] E. Eriksson, T. R. Hansen, and A. Lykke-Olesen, “Movement-based interaction in camera spaces: A conceptual framework”, *Personal and Ubiquitous Computing*, vol. 11, no. 8, pp. 621–632, 2007. DOI: 10.1007/s00779-006-0134-z.

[18] F. Kistler, “Full body interaction: Design, implementation, and user support”, 2016.

[19] A. Sundaresan and R. Chellappa, “Markerless Motion Capture using Multiple Cameras”, in *Computer Vision for Interactive and Intelligent Environment (CVIIE’05)*, Nov. 2005, pp. 15–26. DOI: 10.1109/CVIIE.2005.13.

[20] D. Michelis and J. Müller, “The audience funnel: Observations of gesture based interaction with multiple large displays in a city center”, *Intl. Journal of Human–Computer Interaction*, vol. 27, no. 6, pp. 562–579, 2011. DOI: 10.1080/10447318.2011.555299.

[21] Y. Rogers and H. Muller, “A framework for designing sensor-based interactions to promote exploration and reflection in play”, *International Journal of Human-Computer Studies*, vol. 64, no. 1, pp. 1–14, Jan. 2006, ISSN: 1071-5819. DOI: 10.1016/j.ijhcs.2005.05.004.

[22] H. Joo, T. Simon, X. Li, *et al.*, “Panoptic studio: A massively multiview system for social interaction capture”, *IEEE transactions on pattern analysis and machine intelligence*, vol. 41, no. 1, pp. 190–204, 2017.

[23] A. Sridhar and A. Sowmya, “Multiple camera, multiple person tracking with pointing gesture recognition in immersive environments”, in *International Symposium on Visual Computing*, Springer, 2008, pp. 508–519.

[24] M. Akten, *Body Paint performance at "Clicks or Mortar"*, Jun. 2009. [Online]. Available: https://vimeo.com/5094780.

[25] T. Schofield, J. Vines, T. Higham, E. Carter, M. Atken, and A. Golding, “Trigger shift: Participatory design of an augmented theatrical performance with young people”, in *Proceedings of the 9th ACM Conference on Creativity & Cognition*, Sydney Australia: ACM, Jun. 2013, pp. 203–212, ISBN: 978-1-4503-2150-1. DOI: 10.1145/2466627.2466640.

[26] D. Fallman, "Design-oriented human-computer interaction", in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '03, New York, NY, USA: Association for Computing Machinery, Apr. 2003, pp. 225–232, ISBN: 978-1-58113-630-2. DOI: 10.1145/642611.642652.

[27] P. Sundström, A. Taylor, K. Grufberg, N. Wirström, J. Solsona Belenguer, and M. Lundén, "Inspirational bits: Towards a shared understanding of the digital material", in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '11, New York, NY, USA: Association for Computing Machinery, May 2011, pp. 1561–1570, ISBN: 978-1-4503-0228-9. DOI: 10.1145/1978942.1979170.

[28] H. Sørum, "Design of Digital Products in the Future: A Study of Interaction Design Students and Their Perceptions on Design Issues", in *Design, User Experience, and Usability: Theory, Methodology, and Management*, A. Marcus and W. Wang, Eds., ser. Lecture Notes in Computer Science, Cham: Springer International Publishing, 2017, pp. 740–754, ISBN: 978-3-319-58634-2. DOI: 10.1007/978-3-319-58634-2_53.

[29] O. Alemi, P. Pasquier, and C. Shaw, "Mova: Interactive Movement Analytics Platform", in *Proceedings of the 2014 International Workshop on Movement and Computing - MOCO '14*, Paris, France: ACM Press, 2014, pp. 37–42, ISBN: 978-1-4503-2814-2. DOI: 10.1145/2617995.2618002.

[30] S. Amin and J. Burke, "OpenMoves: A System for Interpreting Person-Tracking Data", in *Proceedings of the 5th International Conference on Movement and Computing*, ACM, 2018, p. 13, ISBN: 1-4503-6504-3.

[31] L. Liu, D. Long, and B. Magerko, "MoViz: A Visualization Tool for Comparing Motion Capture Data Clustering Algorithms", in *Proceedings of the 7th International Conference on Movement and Computing*, Jersey City/Virtual NJ USA: ACM, Jul. 2020, pp. 1–8, ISBN: 978-1-4503-7505-4. DOI: 10.1145/3401956.3404228.

[32] L. Dahl and F. Visi, "Modosc: A Library of Real-Time Movement Descriptors for Marker-Based Motion Capture", in *Proceedings of the 5th International Conference on Movement and Computing*, ser. MOCO '18, New York, NY, USA: Association for Computing Machinery, Jun. 2018, pp. 1–4, ISBN: 978-1-4503-6504-8. DOI: 10.1145/3212721.3212842.

[33] B. J. Oates, *Researching information systems and computing*. Sage, 2005.

[34] J. Grudin, "Three faces of human-computer interaction", *IEEE Annals of the History of Computing*, vol. 27, no. 4, pp. 46–62, 2005.

[35] W. Liu, "Natural user interface-next mainstream product user interface", in *2010 IEEE 11th International Conference on Computer-Aided Industrial Design & Conceptual Design 1*, vol. 1, IEEE, 2010, pp. 203–205, ISBN: 1-4244-7974-6.

[36] D. Wigdor and D. Wixon, *Brave NUI world: designing natural user interfaces for touch and gesture*. Elsevier, 2011, ISBN: 0-12-382232-7.

[37] S. K. Bailey and C. I. Johnson, "Performance on a natural user interface task is correlated with higher gesture production", in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 63, SAGE Publications Sage CA: Los Angeles, CA, 2019, pp. 1384–1388.

[38] S.-M. Wang, S.-X. Lin, and C. Huang, "Smart TV tangible interaction natural user interface design in the scenario of future classroom", in *2016 IEEE International Conference on Consumer Electronics-Taiwan (ICCE-TW)*, IEEE, 2016, pp. 1–2, ISBN: 1-5090-2073-X.

[39] E. Chatzidaki, M. Xenos, and C. Machaira, "A natural user interface game for the evaluation of children with learning difficulties.", in *WOCCI*, 2017, pp. 17–22.

[40] E. Christinaki, N. Vidakis, and G. Triantafyllidis, "Facial expression recognition teaching to preschoolers with autism: A natural user interface approach", in *Proceedings of the 6th Balkan Conference in Informatics*, ACM, 2013, pp. 141–148, ISBN: 1-4503-1851-7.

[41] R. Gündogdu, A. Bejan, C. Kunze, and M. Wölfel, "Activating people with dementia using natural user interface interaction on a surface computer", in *Proceedings of the 11th EAI International Conference on Pervasive Computing Technologies for Healthcare*, ACM, 2017, pp. 386–394.

[42] R. N. Madeira, L. Costa, and O. Postolache, "Physiomate-pervasive physical rehabilitation based on nui and gamification", in *2014 International Conference and Exposition on Electrical and Power Engineering (EPE)*, IEEE, 2014, pp. 612–616.

[43] P. A. Rego, P. M. Moreira, and L. P. Reis, "Natural user interfaces in serious games for rehabilitation", in *6th Iberian Conference on Information Systems and Technologies (CISTI 2011)*, IEEE, 2011, pp. 1–4, ISBN: 989-96247-5-6.

[44] B. Loureiro and R. Rodrigues, "Multi-touch as a natural user interface for elders: A survey", in *6th Iberian Conference on Information Systems and Technologies (CISTI 2011)*, IEEE, 2011, pp. 1–6, ISBN: 989-96247-5-6.

[45] R. Lun and W. Zhao, "A survey of applications and human motion recognition with microsoft kinect", *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 29, no. 5, p. 1 555 008, 2015.

[46] S. Subramanian, N. Skjæret-Maroni, and Y. Dahl, "Systematic review of design guidelines for full-body interactive games", *Interacting with Computers*, vol. 32, no. 4, pp. 367–406, 2020.

[47] E. Márquez Segura, A. Waern, J. Moen, and C. Johansson, "The design space of body games: Technological, physical, and social design", in *Proceedings of the SIGCHI conference on Human Factors in computing systems*, 2013, pp. 3365–3374.

[48] C. Hummels, K. C. J. Overbeeke, and S. Klooster, "Move to get moved: A search for methods, tools and knowledge to design for expressive and rich movement-based interaction", *Personal and Ubiquitous Computing*, vol. 11, no. 8, pp. 677–690, Dec. 2007, ISSN: 1617-4917. DOI: 10.1007/s00779-006-0135-y.

[49] D. Wilde, T. Schiphorst, and S. Klooster, "Move to design/design to move: A conversation about designing for the body", *Interactions*, vol. 18, no. 4, pp. 22–27, Jul. 2011, ISSN: 1072-5520. DOI: 10.1145/1978822.1978828.

[50] A. B. Godbehere, A. Matsukawa, and K. Goldberg, "Visual tracking of human visitors under variable-lighting conditions for a responsive audio art installation", in *2012 American Control Conference (ACC)*, IEEE, 2012, pp. 4305–4312.

[51] L. Loke and T. Robertson, "Design representations of moving bodies for interactive, motion-sensing spaces", *International Journal of Human-Computer Studies*, vol. 67, no. 4, pp. 394–410, 2009.

[52] L. Malinverni, E. Ackermann, and N. Pares, "Experience as an object to think with: From sensing-in-action to making-sense of action in full-body interaction learning environments", in *Proceedings of the TEI'16: Tenth International Conference on Tangible, Embedded, and Embodied Interaction*, 2016, pp. 332–339.

[53] J. Wainwright, "Random bodies: Praxis: Random international", in *Digital Fabrication in Interior Design*, Routledge, 2021, pp. 56–65.

[54] S. U. Ahmed, "Interaction and interactivity: In the context of digital interactive art installation", in *Human-Computer Interaction. Interaction in Context: 20th International Conference, HCI International 2018, Las Vegas, NV, USA, July 15–20, 2018, Proceedings, Part II 20*, Springer, 2018, pp. 241–257.

[55] A. T. Lam, J. Ma, C. Barr, *et al.*, "First-hand, immersive full-body experiences with living cells through interactive museum exhibits", *Nature biotechnology*, vol. 37, no. 10, pp. 1238–1241, 2019.

[56] S. Mishra and F. Cafaro, "Full body interaction beyond fun: Engaging museum visitors in human-data interaction", in *Proceedings of the Twelfth International Conference on Tangible, Embedded, and Embodied Interaction*, 2018, pp. 313–319.

[57] L. Malinverni and N. P. Burguès, "The medium matters: The impact of full-body interaction on the socio-affective aspects of collaboration", in *Proceedings of the 14th International Conference on Interaction Design and Children*, 2015, pp. 89–98.

[58] M.-M. Schaper and N. Pares, "Making sense of body and space through full-body interaction design: A case study", in *Proceedings of the The 15th International Conference on Interaction Design and Children*, 2016, pp. 613–618.

[59] F. Cafaro, L. Lyons, and A. N. Antle, "Framed guessability: Improving the discoverability of gestures and body movements for full-body interaction", in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 2018, pp. 1–12.

[60] J. Müller, R. Walter, G. Bailly, M. Nischt, and F. Alt, "Looking glass: A field study on noticing interactivity of a shop window", in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2012, pp. 297–306.

[61] B. Kules, H. Kang, C. Plaisant, A. Rose, and B. Shneiderman, "Immediate usability: A case study of public access design for a community photo library", *Interacting with Computers*, vol. 16, no. 6, pp. 1171–1193, 2004.

[62] C. Sandifer, "Time-based behaviors at an interactive science museum: Exploring the differences between weekday/weekend and family/nonfamily visitors", *Science Education*, vol. 81, no. 6, pp. 689–701, 1997.

[63] D. Norman, *The design of everyday things: Revised and expanded edition*. Basic books, 2013.

[64] T. Burns, *Erving goffman*. Routledge, 2002.

[65] S. Benford, H. Schnädelbach, B. Koleva, *et al.*, "Expected, sensed, and desired: A framework for designing sensing-based interaction", *ACM Transactions on Computer-Human Interaction (TOCHI)*, vol. 12, no. 1, pp. 3–30, 2005. DOI: 10.1145/1057237.1057239.

[66] J. Martinez, R. Hossain, J. Romero, and J. J. Little, "A Simple Yet Effective Baseline for 3d Human Pose Estimation", in *2017 IEEE International Conference on Computer Vision (ICCV)*, Venice: IEEE, Oct. 2017, pp. 2659–2668, ISBN: 978-1-5386-1032-9. DOI: 10.1109/ICCV.2017.288.

[67] F. Moreno-Noguer, "3D Human Pose Estimation From a Single Image via Distance Matrix Regression", in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 2823–2832.

[68] X. Sun, J. Shang, S. Liang, and Y. Wei, "Compositional Human Pose Regression", in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2602–2611.

[69] G. Pavlakos, X. Zhou, and K. Daniilidis, "Ordinal Depth Supervision for 3D Human Pose Estimation", in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 7307–7316.

[70] R. Hartley and A. Zisserman, *Multiple view geometry in computer vision*. Cambridge: Cambridge University Press, 2004, ISBN: 978-0-511-18618-9.

[71] S. Amin, M. Andriluka, M. Rohrbach, and B. Schiele, "Multi-view Pictorial Structures for 3D Human Pose Estimation", in *Procedings of the British Machine Vision Conference 2013*, Bristol: British Machine Vision Association, 2013, pp. 45.1–45.11, ISBN: 978-1-901725-49-0. DOI: 10.5244/C.27.45.

[72] H. Qiu, C. Wang, J. Wang, N. Wang, and W. Zeng, "Cross View Fusion for 3D Human Pose Estimation", in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, Seoul, Korea (South): IEEE, Oct. 2019, pp. 4341–4350, ISBN: 978-1-72814-803-8. DOI: 10.1109/ICCV.2019.00444.

[73] K. Iskakov, E. Burkov, V. Lempitsky, and Y. Malkov, "Learnable Triangulation of Human Pose", in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, Seoul, Korea (South): IEEE, Oct. 2019, pp. 7717–7726, ISBN: 978-1-72814-803-8. DOI: 10.1109/ICCV.2019.00781.

[74] U. Iqbal, P. Molchanov, and J. Kautz, "Weakly-Supervised 3D Human Pose Learning via Multi-View Images in the Wild", in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 5243–5252.

[75] R. Xie, C. Wang, and Y. Wang, "MetaFuse: A Pre-trained Fusion Model for Human Pose Estimation", in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 13 686–13 695.

[76] H. Tu, C. Wang, and W. Zeng, "VoxelPose: Towards Multi-camera 3D Human Pose Estimation in Wild Environment", in *Computer Vision – ECCV 2020*, A. Vedaldi, H. Bischof, T. Brox, and J.-M. Frahm, Eds., ser. Lecture Notes in Computer Science, Cham: Springer International Publishing, 2020, pp. 197–212, ISBN: 978-3-030-58452-8. DOI: 10.1007/978-3-030-58452-8_12.

[77] J. Lin and G. H. Lee, "Multi-view multi-person 3d pose estimation with plane sweep stereo", in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 11 886–11 895.

[78] H. Ye, W. Zhu, C. Wang, R. Wu, and Y. Wang, "Faster voxelpose: Real-time 3d human pose estimation by orthographic projection", in *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part VI*, Springer, 2022, pp. 142–159.

[79] V. Belagiannis, S. Amin, M. Andriluka, B. Schiele, N. Navab, and S. Ilic, "3D Pictorial Structures for Multiple Human Pose Estimation", in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 1669–1676.

[80] V. Belagiannis, S. Amin, M. Andriluka, B. Schiele, N. Navab, and S. Ilic, "3D Pictorial Structures Revisited: Multiple Human Pose Estimation", *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 10, pp. 1929–1942, Oct. 2016, ISSN: 0162-8828, 2160-9292, 1939-3539. DOI: 10.1109/TPAMI.2015.2509986.

[81] S. Ershadi-Nasab, E. Noury, S. Kasaei, and E. Sanaei, "Multiple human 3D pose estimation from multiview images", *Multimed Tools Appl*, vol. 77, no. 12, pp. 15 573–15 601, Jun. 2018, ISSN: 1573-7721. DOI: 10.1007/s11042-017-5133-8.

[82] J. Dong, W. Jiang, Q. Huang, H. Bao, and X. Zhou, "Fast and Robust Multi-Person 3D Pose Estimation From Multiple Views", in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, CA, USA: IEEE, Jun. 2019, pp. 7784–7793, ISBN: 978-1-72813-293-8. DOI: 10.1109/CVPR.2019.00798.

[83] H. Chen, P. Guo, P. Li, G. H. Lee, and G. Chirikjian, "Multi-person 3D Pose Estimation in Crowded Scenes Based on Multi-view Geometry", in *Computer Vision – ECCV 2020*, A. Vedaldi, H. Bischof, T. Brox, and J.-M. Frahm, Eds., ser. Lecture Notes in Computer Science, Cham: Springer International Publishing, 2020, pp. 541–557, ISBN: 978-3-030-58580-8. DOI: 10.1007/978-3-030-58580-8_32.

[84] A. Kadkhodamohammadi and N. Padoy, "A generalizable approach for multi-view 3D human pose regression", *Machine Vision and Applications*, vol. 32, no. 1, p. 6, Oct. 2020, ISSN: 1432-1769. DOI: 10.1007/s00138-020-01120-2.

[85] Y. Zhang, L. An, T. Yu, X. Li, K. Li, and Y. Liu, "4D Association Graph for Realtime Multi-Person Motion Capture Using Multiple Video Cameras", in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, WA, USA: IEEE, Jun. 2020, pp. 1321–1330, ISBN: 978-1-72817-168-5. DOI: 10.1109/CVPR42600.2020.00140.

[86] Y. Tao, A. Both, R. I. Silveira, *et al.*, "A comparative analysis of trajectory similarity measures", *GIScience & Remote Sensing*, vol. 58, no. 5, pp. 643–669, Jul. 2021, ISSN: 1548-1603. DOI: 10.1080/15481603.2021.1908927.

[87] F. Noroozi, C. A. Corneanu, D. Kamińska, T. Sapiński, S. Escalera, and G. Anbarjafari, "Survey on Emotional Body Gesture Recognition", *IEEE Transactions on Affective Computing*, vol. 12, no. 2, pp. 505–523, Apr. 2021, Conference Name: IEEE Transactions on Affective Computing, ISSN: 1949-3045. DOI: 10.1109/TAFFC.2018.2874986.

[88] S. F. Alaoui, F. Bevilacqua, and C. Jacquemin, "Interactive Visuals as Metaphors for Dance Movement Qualities", *ACM Trans. Interact. Intell. Syst.*, vol. 5, no. 3, 13:1–13:24, Sep. 2015, ISSN: 2160-6455. DOI: 10.1145/2738219.

[89] P. Laube, *Computational Movement Analysis* (SpringerBriefs in Computer Science). Cham: Springer International Publishing, 2014, ISBN: 978-3-319-10267-2. DOI: 10.1007/978-3-319-10268-9.

[90] M. C. González, C. A. Hidalgo, and A.-L. Barabási, "Understanding individual human mobility patterns", *Nature*, vol. 453, no. 7196, pp. 779–782, Jun. 2008, ISSN: 1476-4687. DOI: 10.1038/nature06958.

[91] M. Andersson, J. Gudmundsson, P. Laube, and T. Wolle, "Reporting Leaders and Followers among Trajectories of Moving Point Objects", *Geoinformatica*, vol. 12, no. 4, pp. 497–528, Dec. 2008, ISSN: 1573-7624. DOI: 10.1007/s10707-007-0037-9.

[92] S. Gong, J. Cartlidge, R. Bai, Y. Yue, Q. Li, and G. Qiu, "Extracting activity patterns from taxi trajectory data: A two-layer framework using spatio-temporal clustering, Bayesian probability and Monte Carlo simulation", *International Journal of Geographical Information Science*, vol. 34, no. 6, pp. 1210–1234, Jun. 2020, Publisher: Taylor & Francis, ISSN: 1365-8816. DOI: 10.1080/13658816.2019.1641715.

[93] M. Konzack, T. McKetterick, T. Ophelders, *et al.*, "Visual analytics of delays and interaction in movement data", *International Journal of Geographical Information Science*, vol. 31, no. 2, pp. 320–345, Feb. 2017, ISSN: 1365-8816, 1362-3087. DOI: 10.1080/13658816.2016.1199806.

[94] P. Ranacher and K. Tzavella, "How to compare movement? A review of physical movement similarity measures in geographic information science and beyond", *Cartography and Geographic Information Science*, vol. 41, no. 3, pp. 286–307, May 2014, ISSN: 1523-0406, 1545-0465. DOI: 10.1080/15230406.2014.890071.

[95] A. Escamilla, J. Melenchón, C. Monzo, and J. A. Morán, "Interaction designers' perceptions of using motion-based full-body features", *International Journal of Human-Computer Studies*, vol. 155, p. 102 697, Nov. 2021, ISSN: 10715819. DOI: 10.1016/j.ijhcs.2021.102697.

[96] A. Camurri, S. Hashimoto, M. Ricchetti, *et al.*, "EyesWeb: Toward Gesture and Affect Recognition in Interactive Dance and Music Systems", *Computer Music Journal*, vol. 24, no. 1, pp. 57–69, 2000, Publisher: The MIT Press, ISSN: 0148-9267.

[97] A. Friberg, *A fuzzy analyzer of emotional expression in music performance and body motion* (Proceedings of Music and Music Science). Jan. 2004, vol. 10.

[98] M. Feldmeier and J. A. Paradiso, "An Interactive Music Environment for Large Groups with Giveaway Wireless Motion Sensors", *Computer Music Journal*, vol. 31, no. 1, pp. 50–67, Mar. 2007, ISSN: 0148-9267, 1531-5169. DOI: 10.1162/comj.2007.31.1.50.

[99] F. Bevilacqua, F. Guédy, N. Schnell, E. Fléty, and N. Leroy, "Wireless sensor interface and gesture-follower for music pedagogy", *New York*, p. 6, 2007.

[100] M. Gillies, "Understanding the Role of Interactive Machine Learning in Movement Interaction Design", *ACM Transactions on Computer-Human Interaction*, vol. 26, no. 1, 5:1–5:34, Feb. 2019, ISSN: 1073-0516. DOI: 10.1145/3287307.

[101] N. Plant, R. Gibson, C. G. Diaz, *et al.*, "Movement interaction design for immersive media using interactive machine learning", in *Proceedings of the 7th International Conference on Movement and Computing*, ser. MOCO '20, New York, NY, USA: Association for Computing Machinery, Jul. 2020, pp. 1–2, ISBN: 978-1-4503-7505-4. DOI: 10.1145/3401956.3404252.

[102] R. Fiebrink, P. R. Cook, and D. Trueman, "Human model evaluation in interactive supervised learning", in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Vancouver BC Canada: ACM, May 2011, pp. 147–156, ISBN: 978-1-4503-0228-9. DOI: 10.1145/1978942.1978965.

[103] J. Françoise, Y. Candau, S. Fdili Alaoui, and T. Schiphorst, "Designing for Kinesthetic Awareness: Revealing User Experiences through Second-Person Inquiry", in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, Denver Colorado USA: ACM, May 2017, pp. 5171–5183, ISBN: 978-1-4503-4655-9. DOI: 10.1145/3025453.3025714.

[104] B. Caramiaux, A. Altavilla, S. G. Pobiner, and A. Tanaka, "Form Follows Sound: Designing Interactions from Sonic Memories", in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, Seoul Republic of Korea: ACM, Apr. 2015, pp. 3943–3952, ISBN: 978-1-4503-3145-6. DOI: 10.1145/2702123.2702515.

[105] M. Munaro, F. Basso, and E. Menegatti, "OpenPTrack: Open source multi-camera calibration and people tracking for RGB-D camera networks", *Robotics and Autonomous Systems*, vol. 75, pp. 525–538, Jan. 2016, ISSN: 0921-8890. DOI: 10.1016/j.robot.2015.10.004.

[106] J. Maccallum, R. Gottfried, I. Rostovtsev, J. Bresson, and A. Freed, "Dynamic Message-Oriented Middleware with Open Sound Control and Odot", in *International Computer Music Conference*, 2015.

[107] C. Hilton, N. Plant, C. González Díaz, *et al.*, "InteractML: Making machine learning accessible for creative practitioners working with movement interaction in immersive media", in *Proceedings of the 27th ACM Symposium on Virtual Reality Software and Technology*, Osaka Japan: ACM, Dec. 2021, pp. 1–10, ISBN: 978-1-4503-9092-7. DOI: 10.1145/3489849.3489879.

[108] J. Bernard, A. Vögele, R. Klein, and D. Fellner, "Approaches and Challenges in the Visual-interactive Comparison of Human Motion Data:" in *Proceedings of the 12th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*, Porto, Portugal: SCITEPRESS - Science and Technology Publications, 2017, pp. 217–224, ISBN: 978-989-758-228-8. DOI: 10.5220/0006127502170224.

[109] V. Arpatzoglou, A. Kardara, A. Diehl, B. Flueckiger, S. Helmer, and R. Pajarola, "DanceMoves: A Visual Analytics Tool for Dance Movement Analysis", *EuroVis 2021 - Short Papers*, 5 pages, 2021. DOI: 10.2312/EVS.20211057.

[110] D. Long, L. Liu, S. Gujrania, C. Naomi, and B. Magerko, "Visualizing improvisation in luminai, an ai partner for co-creative dance", in *Proceedings of the 7th International Conference on Movement and Computing*, ser. MOCO '20, Jersey City/Virtual, NJ, USA: Association for Computing Machinery, 2020. DOI: 10.1145/3401956.3404258.

[111] A. Nili, M. Tate, and D. Johnstone, "A framework and approach for analysis of focus group data in information systems research", *Communications of the Association for Information Systems*, vol. 40, 1:1–1:21, 2017. DOI: 10.17705/1CAIS.04001.

[112] P. Hernández Salazar, *Métodos cualitativos para estudiar a los usuarios de la información.* México: UNAM, Centro Universitario de Investigaciones Bibliotecológicas, 2008., 2008.

[113] H. O'Connor and C. Madge, "Focus groups in cyberspace: Using the Internet for qualitative research", *Qualitative Market Research: An International Journal*, vol. 6, no. 2, pp. 133–143, Jan. 2003, ISSN: 1352-2752. DOI: 10.1108/13522750310470190.

[114] M. M. Montoya-Weiss, A. P. Massey, and D. L. Clapper, "On-line focus groups: Conceptual issues and a research tool", *European Journal of Marketing*, vol. 32, no. 7/8, pp. 713–723, Jan. 1998. DOI: 10.1108/03090569810224100.

[115] D. J. Reid and F. J. Reid, "Online focus groups: An in-depth comparison of computer-mediated and conventional focus group discussions", *International journal of market research*, vol. 47, no. 2, pp. 131–162, 2005, Publisher: SAGE Publications Sage UK: London, England.

[116] J. Lazar, J. H. Feng, and H. Hochheiser, *Research methods in human-computer interaction*. Morgan Kaufmann, 2017.

[117] E. Vaportzis, M. Giatsi Clausen, and A. J. Gow, "Older Adults Perceptions of Technology and Barriers to Interacting with Tablet Computers: A Focus Group Study", *Frontiers in Psychology*, vol. 8, 2017, ISSN: 1664-1078. DOI: 10.3389/fpsyg.2017.01687.

[118] J. A. Hall and M. L. Knapp, *Nonverbal Communication*. Walter de Gruyter, Jan. 2013, ISBN: 978-3-11-023815-0.

[119] A. J. Onwuegbuzie, N. L. Leech, and K. M. T. Collins, "Innovative Data Collection Strategies in Qualitative Research", *Qualitative Report*, vol. 15, no. 3, pp. 696–726, May 2010, ISSN: 1052-0147.

[120] M. Eriksson, "Using Social Science in Design", in *User Centric Media*, P. Daras and O. M. Ibarra, Eds., ser. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, Berlin, Heidelberg: Springer, 2010, pp. 361–365, ISBN: 978-3-642-12630-7. DOI: 10.1007/978-3-642-12630-7_46.

[121] G. N. Vilaza and J. E. Bardram, "Sharing Access to Behavioural and Personal Health Data: Designers' Perspectives on Opportunities and Barriers", in *Proceedings of the 13th EAI International Conference on Pervasive Computing Technologies for Healthcare*, ser. PervasiveHealth'19, New York, NY, USA: Association for Computing Machinery, May 2019, pp. 346–350, ISBN: 978-1-4503-6126-2. DOI: 10.1145/3329189.3329229.

[122] H. Seifi, M. Chun, C. Gallacher, O. Schneider, and K. E. MacLean, "How Do Novice Hapticians Design? A Case Study in Creating Haptic Learning Environments", *IEEE Transactions on Haptics*, vol. 13, no. 4, pp. 791–805, Oct. 2020, ISSN: 2329-4051. DOI: 10.1109/TOH.2020.2968903.

[123] I. A. Owusu, J. J. Daalhuizen, and P. J. Stappers, "Flexibility of choice and perceived impact of using design methods", in *DS 70: Proceedings of DESIGN 2012, the 12th International Design Conference, Dubrovnik, Croatia*, 2012, pp. 1061–1070.

[124] F. K. Ozenc, M. Kim, J. Zimmerman, S. Oney, and B. Myers, "How to support designers in getting hold of the immaterial material of software", in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '10, New York, NY, USA: Association for Computing Machinery, Apr. 2010, pp. 2513–2522, ISBN: 978-1-60558-929-9. DOI: 10.1145/1753326.1753707.

[125] K. O'hara, R. Harper, H. Mentis, A. Sellen, and A. Taylor, "On the naturalness of touchless: Putting the "interaction" back into nui", *ACM Transactions on Computer-Human Interaction*, vol. 20, no. 1, 5:1–5:25, Apr. 2013, ISSN: 1073-0516. DOI: 10.1145/2442106.2442111.

[126] E. Márquez Segura, L. Turmo Vidal, A. Rostami, and A. Waern, "Embodied Sketching", in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, ser. CHI '16, New York, NY, USA: Association for Computing Machinery, May 2016, pp. 6014–6027, ISBN: 978-1-4503-3362-7. DOI: 10.1145/2858036.2858486.

[127] S. Garrido-Jurado, R. Muñoz-Salinas, F. J. Madrid-Cuevas, and M. J. Marín-Jiménez, "Automatic generation and detection of highly reliable fiducial markers under occlusion", *Pattern Recognition*, vol. 47, no. 6, pp. 2280–2292, Jun. 2014, ISSN: 0031-3203. DOI: 10.1016/j.patcog.2014.01.005.

[128] H.-S. Fang, J. Li, H. Tang, *et al.*, "AlphaPose: Whole-Body Regional Multi-Person Pose Estimation and Tracking in Real-Time", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Nov. 2022.

[129] J. Redmon and A. Farhadi, *YOLOv3: An Incremental Improvement*, Apr. 2018. DOI: 10.48550/arXiv.1804.02767.

[130] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition", in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA: IEEE, Jun. 2016, pp. 770–778, ISBN: 978-1-4673-8851-1. DOI: 10.1109/CVPR.2016.90.

[131] T.-Y. Lin, M. Maire, S. Belongie, *et al.*, "Microsoft COCO: Common Objects in Context", in *Computer Vision – ECCV 2014*, D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, Eds., ser. Lecture Notes in Computer Science, Cham: Springer International Publishing, 2014, pp. 740–755, ISBN: 978-3-319-10602-1. DOI: 10.1007/978-3-319-10602-1_48.

[132] L. Chen, H. Ai, Z. Zhuang, and C. Shang, "Real-time Multiple People Tracking with Deeply Learned Candidate Selection and Person Re-Identification", *2018 IEEE International Conference on Multimedia and Expo (ICME)*, pp. 1–6, Jul. 2018. DOI: 10.1109/ICME.2018.8486597.

[133] J. Tanke and J. Gall, "Iterative Greedy Matching for 3D Human Pose Tracking from Multiple Views", in *Pattern Recognition*, G. A. Fink, S. Frintrop, and X. Jiang, Eds., vol. 11824, Series Title: Lecture Notes in Computer Science, Cham: Springer International Publishing, 2019, pp. 537–550, ISBN: 978-3-030-33675-2. DOI: 10.1007/978-3-030-33676-9_38.

[134] D. Glowinski, N. Dael, A. Camurri, G. Volpe, M. Mortillaro, and K. Scherer, "Toward a Minimal Representation of Affective Gestures", *IEEE Transactions on Affective Computing*, vol. 2, no. 2, pp. 106–118, Apr. 2011, Conference Name: IEEE Transactions on Affective Computing, ISSN: 1949-3045. DOI: 10.1109/T-AFFC.2011.7.

[135] D. Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 5, pp. 603–619, May 2002, Conference Name: IEEE Transactions on Pattern Analysis and Machine Intelligence, ISSN: 1939-3539. DOI: 10.1109/34.1000236.

[136] M. Ester, H.-P. Kriegel, and X. Xu, "A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise", in *kdd*, vol. 96, 1996, pp. 226–231.

[137] C. Goodall, "Procrustes Methods in the Statistical Analysis of Shape", *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 53, no. 2, pp. 285–321, 1991, ISSN: 2517-6161. DOI: 10.1111/j.2517-6161.1991.tb01825.x.

[138] E. M. Barhorst-Cates, M. W. Isaacs, L. J. Buxbaum, and A. L. Wong, "Action imitation via trajectory-based or posture-based planning", *Human Movement Science*, vol. 83, p. 102 951, Jun. 2022, ISSN: 0167-9457. DOI: 10.1016/j.humov.2022.102951.

[139] E. R. Tufte, "The Visual Display of Quantitative Information", *The Journal for Healthcare Quality (JHQ)*, vol. 7, no. 3, p. 15, Jul. 1985, ISSN: 1062-2551.

[140] D. Chi, M. Costa, L. Zhao, and N. Badler, "The EMOTE model for effort and shape", in *Proceedings of the 27th annual conference on Computer graphics and interactive techniques - SIGGRAPH '00*, Not Known: ACM Press, 2000, pp. 173–182, ISBN: 978-1-58113-208-3. DOI: 10.1145/344779.352172.

[141] D. Swaminathan, H. Thornburg, J. Mumford, *et al.*, "A Dynamic Bayesian Approach to Computational Laban Shape Quality Analysis", *Advances in Human-Computer Interaction*, vol. 2009, e362651, May 2009, Publisher: Hindawi, ISSN: 1687-5893. DOI: 10.1155/2009/362651.

[142] T. Schiphorst, "Soft(n): Toward a somaesthetics of touch", in *CHI '09 Extended Abstracts on Human Factors in Computing Systems*, Boston MA USA: ACM, Apr. 2009, pp. 2427–2438, ISBN: 978-1-60558-247-4. DOI: 10.1145/1520340.1520345.

[143] P. Subyen, D. S. Maranan, K. Calrson, T. Schiphorst, and P. Pasquier, "Flow: Expressing movement quality", in *Learning, The User in Flux Workshop at CHI*, vol. 11, 2011.

[144] D. Silang Maranan, S. Fdili Alaoui, T. Schiphorst, P. Pasquier, P. Subyen, and L. Bartram, "Designing for movement: Evaluating computational models using LMA effort qualities", in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '14, New York, NY, USA: Association for Computing Machinery, Apr. 2014, pp. 991–1000, ISBN: 978-1-4503-2473-1. DOI: 10.1145/2556288.2557251.

[145] H. M. Mentis and C. Johansson, "Seeing movement qualities", in *Proceedings of the sigchi conference on human factors in computing systems*, 2013, pp. 3375–3384.