REAL MATERIAL ETHEREAL

Proceedings of the Annual Design Research Conference 2019

Edited by Laura Harper

Published in Melbourne Australia, by Monash University, 2020
ISBN: 978-1-921994-52-4

The bibliographic citation for this paper is:

Charlotte Firth, Kate Dunn and Matthias Hank Haeusler. "Motion capture techniques and discussion of role in developing complex robotic systems." In *Proceedings of the Annual Design Research Conference 2019: Real/Material/Ethereal,* edited by Laura Harper, 192-200. Melbourne: Monash University, 2020

Motion capture techniques and discussion of role in developing complex robotic systems

Abstract.

This paper investigates the use of motion capture techniques and software to track human movement and their potential role in developing complex robotic systems for the construction industry. As the construction industry faces adversity from increased demands from rapid population growth, due to the decline in skilled trades workers and heavy reliance on traditional fabrication techniques that are physically demanding. The results of these challenges potentially lead to the escalation of housing costs and construction time. An innovative approach was proposed to prevent this cascade effect, involving the design of a unique, bespoke and interactive relationship between human artisan and machine intelligence via Cyber-Physical Systems and Social Robotics. This review looks at how robots could be implemented onto a construction site to collaborate with workers to reduce unsafe work environments and improve the health of workers and reduce fatigue from repetitive motions.

In previously published work, we established a novel methodology for workflow capture and analysis of carpentry tasks for human-robot collaborative 'actions'. Empirical observations and analysis of the body movements of carpenters was collected and used to aid discussion on how robots can be implemented (*Reinhardt, Dagmar et al. 2019). Building on these findings, this paper discusses how these 'actions' can be captured using motion capturing of arm, hand and finger movements. The research documents which methods and technologies are currently used and discusses the advantages and disadvantages. The data collected from motion capture can be transformed into a training data set for a neural network, which enables a machine to learn the captured movements. Arguing that once the machine learning has been established, all movements can be replicated via a robot arm. This investigation, its proposed hypothesis, methodology, implications, significance and evaluation are presented in the paper.*

Introduction and Research Motivation

The Architecture, Engineering and Construction (AEC) industry is one of the largest industries in the world, with an estimated worth of $10.5 trillion by 2023. However, the future of this industry is at risk as it heavily relies on non-digitised fabrication techniques and a limited pool of skilled trades people, generating a $1.6 trillion-dollar productivity gap within the global economy (McKinsey & Company 2017). This lag in technological uptake and shortfall of skilled trades affects the efficiency and safety of building sites as pressure and demand on the existing workers increases the risk fatigue and injuries. Adding to demand is a steadily growing population which rises by 1.6 % each year and will continue, thereby putting pressure on the development of housing and commercial spaces. Commercial spaces were estimated to increase by 9.3 % in 2018 alone. The construction industry is the third largest industry within Australia and large portion of the Australian economy. To meet demand and maintain a safe working environment it is vital for the construction industry to adopt advances in digital technology and fabrication as a tool for productivity.

The development of robots for the construction industry has been garnering interest since the late 1970's with Japan testing a range of different robots such as hazardous environment robots that could spray fireproofing, but this testing came arguably to a halt when Japan's economy faced turmoil in the 1990's. Another reason for this lack of implementation of robots on construction sites is due to the focus on the design of industrial robots intended for manufacturing rather than the construction industry (Rosenberg et al. 2015).

While industrial robots have been successful in the automation of the manufacturing and automotive industries executing repetitive processes, they are primarily programmed for singular tasks that require extensive programing and accuracy. There is considerable opportunity for a collaborative approach between industry and academia to advance research in this area to meet the needs of the construction industry. To date industrial robots such has KUKA and ABB, have been repurposed to be used in academic architecture and design research in institutions around the globe.

In comparison to industrial robots with their Work Health and Safety restrictions, collaborative robots offer the potential to work on construction sites as they designed to work alongside humans independent of a working cell. Their lightweight structure and intuitive force-torque sensors allow the robot to stop when coming in contact with humans, essential for collaborative work (Rosenberg et al. 2015) Collaborative robots have been developed through the 4th industrial revolution coined as industry 4.0.

Industry 4.0 relies on a set of design principles that include: Interoperability, information transparency, technical assistance and decentralised decisions. These design principles have allowed the manufacturing industry to become not only more efficient but intelligent, allowing for cyber-physical systems to make decisions and act on those decisions autonomously as they are continually monitoring the physical world. Considering the forever changing nature of a construction site and the need to move towards digitalisation, adopting these design principles could be vital in taking the first step. Intelligent cyber-physical systems could be trained to sense a change in not only the physical environment but the changes in human motion and behaviour. There is a growing body of research into 'training' versus 'programming' robots, in particular training of collaborative robots via machine learning and AI (Stanton et al. 2012, Domingos 2012, Levine et al. 2018). Hence this research project took the stand that due to the forever changing nature of a construction site, 'training' of a robot using technologies such as motion capture should be favoured over the 'programming' of a robot.

Research Context - Co Built Research Project

The findings discussed and presented in this paper are part of an extensive research investigation. Developed as joint research between Landcom NSW with Sydney and Western Sydney University, the larger research project aims to address the discrepancy between an unprecedented boom in urban growth combined with a severe shortage in skilled construction labour, while promoting active community engagement in redesigning construction and fabrication methodologies for cities. In order to achieve this goal, we observed how carpenters do their work and analyse their current workflows for potential points where insertion of a collaborative robot could be of assistance and improve efficiencies through visual observations and recordings. During the observations of workflows, particular attention was paid to the worker's movements, space they occupied and, and the time it took to conduct repetitive tasks. As discussed in our paper (Reinhardt, Dagmar et al. 2019) we argue as mentioned earlier that through understanding how trades workers carry out tasks on site, in particular, carpenters, we can have an understanding of tasks that can use a robot as a co-worker.

Consequently, breaking down the workflow of tasks via analysing carpentry jobs, looking towards human-robot collaboration, we were able to create a framework what to capture (the action). We needed in the following, an understanding of how to capture (the movements that define the action). As argued earlier, this research aims to train robots' vs program robots and thus requires large amounts of motion capture data when using supervised machine learning, to train a collaborative robot.

Thus the main objective of this research paper is to investigate the potential of motion capture technologies to create data sets for supervised machine learning algorithms. While the research team is aware that one can also avoid the use of a large dataset in machine learning by using reinforcement learning (unsupervised learning) which requires no inputs, just defined rewards and punishments to learn, (Domingos 2012) we believe motion capture critical for both processes. A detailed capturing of data via motion capture is useful for both supervised and unsupervised learning as it can be used as training data for supervised learning as well as giving humans an understanding of how to define learning framework for reinforcement learning to mimic actions through trial and error.

Objectives:

The paper presents a review of current motion capture technologies as well as a review of motion capture in robotic system development in order to evaluate the feasibility of providing datasets for primary supervised machine learning. This research proposes two main objectives to carry out the review process.

1. To identify all current methods and technologies that are used to develop robotic systems to mimic human movement or to carry out tasks that are deemed repetitive or demanding.

2. Evaluate the advantages and disadvantages of the identified methods and investigate how it was used and if it was successful in achieving goals that can be transferred into the construction industry.

Methodology:

This research reviewed a defined range of methods and technologies for motion capture that could be used to aid the transition of collaborative robots on to a construction site, and compare and analysis appropriate methods via a literature review and review of technical specifications of motion capture. Particular focus was given on motion capturing techniques that produce data sets that are suitable for later supervised machine learning to provide suitable information for cyber-physical systems and elements of social understanding. The focus of this paper is not machine learning. Hence due to limitations in the scope of the paper, we exclude a more extended discussion and refer to the papers referenced for deeper engagement.

Review of Current Methods and Technologies in Motion Capturing:

Motion Capture is defined as a method or a technique that enables users to record a movement or patterns of movements digitally and has been used widely across a range of fields to improve processes and the understanding of the human body. Some of the most popular fields being animation, virtual reality, military, sports science and robotics. Over the years' motion capture has progressed from drawings and basic "glow stick" skeletons into accurate 3D models of the human body that can express real facial expressions and detailed hand movements (Kitagawa 2008). Multiple methods can be used to capture human motion, including Optical, Mechanical, Inertial, Magnetic, and Acoustic. While in the field of robotics, the most commonly used is the optical and inertial methods, the paper will present and discuss all mentioned above.

Optical Motion Capture: Most optical systems use a system of cameras ranging from 4 – 32 cameras that link to a computer that controls each capture. The subjects of the motion capture will be covered in a range of circular markers that are either considered passive or active. Passive markers are reflective and do not emit light, whereas an active marker will emit a light source most commonly LED. These markers can be applied to the skin directly on top specialised suits. Active markers are more suited to areas that have natural lighting as they can emit LED light to be identified whereas lighting needs to be in a controlled environment for passive systems as the cameras need to pick up the reflectivity by directing LED lights on the markers. The advantages of this system is the accuracy of the data is within 1mm error margin ideal for robotic applications, the systems are wireless, allowing for a free range of movement, but on the other hand, it only provides position, so the data requires post-processing if the user required an understanding of a range of motions to complete a task. (Kitagawa 2008)

Mechanical Motion Capture: Mechanical Motion capture systems consist of straight rods and potentiometers (variable resistors) that can be used to measure joint angles of a subject directly. This is considered a simple technique for motion capture, which is flexible, can be used outdoors as it is easily portable, ideal for use on a construction site. Although it can be used anywhere due to the nature of the rods, it can restrict movement not allowing for the full range of motion required for specific tasks. Other types of mechanical systems include data gloves and digital armatures. Wireless mechanical systems provide large capture volumes. Data gloves are an appealing method when trying to capture hand and finger movement. It is an electronic glove that is equipped with built-in sensors that can detect the movement of each finger individually in real-time and output the movement data to a computer in real-time. It is difficult to use optical systems due to occlusion issues, especially if other body motions, are required. The benefits of using a device such as, data gloves, allow a deeper understanding of the ways a trades worker would use specific tools, which an understanding would be needed if a collaborative robot is going to use tools found on site.

Inertial Motion Capture: Inertial motion capture systems do not rely on cameras and external markers; instead, this method uses sensors that detect acceleration and rotational velocities from accelerometers and gyroscopes. A disadvantage of this technology is that there is no reference position (Start or end position) it has to be estimated combining the acceleration and velocity although some more advanced suits use force detection via sudden foot acceleration to update the reference position. The most significant advantage of inertial motion capture that it captures accelerations with the accuracy with an error margin of less than 1 degree (Field et al. 2009). Although considering tasks such as the development of house framing the rate of acceleration could provide a training data set that allowed the robot to anticipate the required speed of their actions.

Magnetic Motion Capture: Magnetic or electromagnetic motion capture systems began when sensors were placed on military aircraft helmets so that the position and orientation of their heads could be tracked in real-time. The method utilises around 6-11 sensors being placed on the body to record joint movement, angles and global rotational values via low-frequency magnetic fields. An advantage of being able to place the sensors is that users can adjust arrangements as required. An issue that comes with this time of motion capture system is that they are prone to distorted outputs due to magnetic and electrical interference. (Field et al. 2009).

Acoustic Motion Capture: Acoustic motion capture work by using microphones and ultrasonic transmitters (measures the distance between two objects using sound waves) that are positioned in specific locations on a human body. This method determines the position of the human body via the intensity of acoustic pulses. A significant disadvantage of this particular method is that arrangement of transmitters and microphones can cause self-occlusion and interference from background noises, temperature and wind if it is used outside, making this process quite ineffective if not used in a controlled environment (Field et al. 2009). A construction site has a large range of noises, from machinery to verbal communication, that would interfere with an acoustic system providing false readings. Although a particular task could be recreated within a controlled environment, it is unnecessary when there are other systems available

Review of Motion Capture in Robotic System development:

Motion capture plays a significant role in developing robotic applications as an intelligent robot must know how to interact safely with a human co-partner. The most popular applications that motion capture is used for, includes; comparing the range of motion between robots and humans, learning from demonstration, real-time control of robots using teleoperation and using the data from motion capture to create training data sets for supervised machine learning algorithms. This section of the review will look at a range of different projects that use motion capture as a method that can control a robot to mimic human movement. This demonstrates the necessary use of motion capture and further steps in creating a robotic system that relies on social robotics and cyber-physical systems to develop a safe collaborative environment for humans on a construction site based on the observations from (Reinhardt, Dagmar et al. 2019).

Inertial motion capture was used in the paper, “Motion capture from inertial sensing for Untethered humanoid teleoperation” (Miller et al. 2004) as it suggests that inertial systems are not as complex as optional and magnetic systems due to the fact it relies only on gravitation force and magnetic fields to function. The project aims to control a robot arm in real-time by using the rotational orientation of the subject's arm to translate from seven to five degrees of freedom for the humanoid robot. The project was successful but faced issues trying to synchronise, the humanoid required that its end-effector to match the tele operator suit which had to be done via searching without feedback. Discussed in the paper is the need for "increased information for human-robot interaction" to move out of the lab setting. For robots in construction, there is also a need for more information which could be acquired through motion capture to aid the transition from research to reality. The project did not use the collected values for robot learning from demonstration but provides the platform to do so, considering the construction industry it is essential that the robot learns something from the data and expand the learnt movements.

Successful integration of collaborative robots onto a construction site would allow the robot to have a social platform that is capable of communication and understanding of human gestures or movements. This ability of communication could be achieved through the implementation of a cyber-physical system that could detect voice commands or gestures. "Robots that imitate humans" (Breazeal & Scassellati 2002) uses motion capture techniques (Optical, mechanical and magnetic) in a controlled environment as a solution to train a robot to understand human behaviour. Reiterated in this work is the value of motion capture data, in the case 3D points in space, to feed into supervised machine learning algorithms. The paper states that robots that have a social understanding perform better in a collaborative environment. Motion capture is the necessary first step in training a cyber-physical system through machine learning to understand changes in the real world to include human motion and behaviour. A robot that could recognise gestures or particular motions, would gain the ability to stop motions or actions, change tools or switch tasks without excessive programming due to the cyber-physical system having awareness of more than just physical changes in the environment but changes in human behaviour. This focus on social behaviour is also prevalent in the previous paper (Reinhardt, Dagmar et al. 2019) as it states the verbal and non-verbal nature of collaborative work on a construction site through the observations of carpenters assembling a stud wall further emphasising the importance of motion capture to understand gestures but informs the next step of the collaborative research to focus on verbal communication and social aspects.

Discussed in "Deep learning-based human motion recognition for predictive context-aware human-robot collaboration" (Wang et al. 2018) is the use of a different kind of motion capture not described in current technologies. This approach is different as a subject completes a motion whilst ten consecutive images are taken for each part of the action. Here they are separated into three categories grasping, holding, and assembling. From here, images are fed into a convoluted neural network. The aim of Peng Wang's paper was for robots to gain contextual awareness of the current task in a factory and actively assist human's workers to complete tasks. Within the construction industry, this form of awareness would allow robots to adapt to multiple tasks, whilst predicting actions in real time. The emphasis on contextual awareness again points towards the necessity of a cyber-physical system on-site to constantly be providing feedback to the robotic systems, allowing a robot to be aware of the current task and to actively assist a human co-partner. Considering the fast nature of the construction industry a cyber-physical system that is able to use the understanding of the motion of certain tasks and anticipate the next steps.

Observations from the greater research suggest tasks of holding a stud wall in place, hold the timber in position for nailing in position, measuring or cutting timber (Reinhardt, Dagmar et al. 2019). Peng Wang, states the importance of a robot's ability to anticipate these tasks is crucial for a safe working environment for humans.

Optical Motion capture was used in "Teleoperation of a humanoid robot using full-body motion capture, example movements, and machine learning." (Stanton et al. 2012) to enable real-time teleoperation control of a humanoid robot. The authors add emphasis to the fact motion capture techniques used with machine learning algorithms will allow for the success of the humanoid in unstructured environments, like urban developments. When looking at the construction industry, it is essential that robots can adapt to unstructured environments, as construction sites are always evolving. Initially, the motion capture data that is captured through a suit attached with a data glove is used to understand the range of motion that a human comparison to the humanoid robot. The robot instructs the human to copy its full range of motion, and this data is collected externally for the supervised machine learning algorithm. The output of data collection was used in for each joint which used feedforward neural networks that allowed for kinematic mapping between the human and the humanoid. They were successfully able to control the robot itself in real-time in a novel sense, showing motion capture techniques were advantageous in creating training data for a supervised machine learning algorithm. The disadvantage of the method presented in this paper is that the post-processing of the data would have been extensive, due to each joint requiring its own neural network for learning how to transfer the human motion into a similar robotic movement. Although successful, reinforcement learning could be considered here to eliminate the need for excessive post processing and programming. As argued before the use of motion capture data was useful in determining the framework for reinforcement learning to map the difference in kinematics.

Evaluation & Significance

From reviewing current motion capture technologies within a range of different applications, it is clear that the understanding of human motion is crucial for human-robot collaboration. A significant aspect that was not expected before reviewing literature was the importance of social aspects that motion capturing techniques can capture. Not only was motion capture used to mimic human movements to carry out tasks intelligently, but motion capture has also been used to developed contextual and social awareness, a point we further highlight from the observations in our paper (Reinhardt et al. 2019). The observation of the way that trades workers communicate to collaborate was a significant aspect of completing tasks shown within the quotes "yeah one there and there and on the bottom there. Yeah, that's right. Don't tighten them with a drill. Just get em on hand tight - you know what I mean? Give me the drill". This communication could be translated via a machine intelligent cyber-physical system for a robot to hand over a drill or brace a stud wall. The data collected to recognise gestures or voice commands could be fed into machine learning algorithms used to train a cyber-physical system awareness of how trades people work, verbally and non-verbally, allowing the robot to anticipate and plan its next action. Anticipation is crucial due to the ever-evolving nature of the construction industry and quick feedback between co-workers. This intelligence changes the robot from being a tool to a co-worker. The second observation through the review of past projects is that machine learning algorithms play a huge role in either developing awareness or recreating actions. Majority of the papers that have been presented focused on supervised learning techniques. Supervised learning is not the only technique to use machine learning algorithms, as unsupervised learning can be used without inputs (Domingos 2012). The motion data can be used as the framework to develop reward systems within the algorithm, again proving it is a crucial step in developing a robotic system that is able to address the inconsistency between a boom in urban growth joined with a shortage in skilled construction labour.

Conclusion and Further Work

The advantages and disadvantages of different motion capture methods and how these methods develop intelligent robotics programs via supervised machine learning are reviewed within this paper. A range of previous works published in the referenced papers that used motion capture as a technique to create robotic systems that mimic human movements or carry out a task have been reviewed to explore the techniques further. Based on both parts, we can conclude the following.

The role of motion capture data was reviewed to understand the importance that data plays in developing robotic systems and how successful it was in achieving the goals of the project. It is clear from this review that motion capture is a vital element in developing robot systems purely in understanding the range of motion required to complete tasks. The next step would be to pre-process the data to create a training set for supervised machine learning, but the referenced papers also indicate that this is a time-consuming and resource extensive undertaking. Within the research context, to capture specific movements of a carpenter, let alone various participants hundreds of times, would provide a suitable data set but would also be time-consuming and an extensive resource undertaking as argued previously.

Hence through conducting this research, we propose that supervised machine learning to train robots might be unnecessary due to the complexity and diversity of tasks occurring on-site and the time it would take to capture each of them.

An alternative to supervised machine learning is unsupervised learning, for example, reinforcement learning, which does not need an input to process output as mentioned earlier. This alternative would allow for machine learning algorithms to use trial and error to develop a tooling path suitable to complete a task, skipping the step of collecting motion capture data for a training set, allowing for more outputs in learnt knowledge to create a safe environment for humans and robots to collaborate in the ever-changing environment. Still, when considering the construction industry and the need for intelligent, safe robots, motion capture is essential in understanding the actions and gestures of trades workers on construction sites to develop a successful integration.

References:

Edited Book:

Kitagawa, Midori. 2008. *MoCap for artists: workflow and techniques for motion capture*. Amsterdam: Boston Elsevier/Focal Press.

Report:

McKinsey & Company. 2017. *Reinventing Construction: A Route to Higher Productivity*. McKinsey Global Institute.

Journal:

Reinhardt, Dagmar et al. 2019. "CoBuilt: Towards a novel methodology for workflow capture and analysis of carpentry tasks for human-robot collaboration". Proceeding of the 37th *eCAADe conference* 37 (SIGraDi 23).

Rosenberg, Elliot et al. 2015. "Smart Architecture-Bots & Industry 4.0 Principles for Architecture". Proceeding of the 33rd *eCAADe Conference* 2: 251-259.

Rosenberg, Elliot; Haeusler, M Hank & Koh, Jeffery. 2015. "From Bob the Builder to Baxter the Builder'". *Proceedings of the 20th International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA 2015).*: 85-94.

Stanton, Christopher; Bogdanovych, Anton & Ratanasena, Edward. 2012. "Teleoperation of a humanoid robot using full-body motion capture, example movements, and machine learning". *Proceedings of Australasian Conference on Robotics and Automation.*

Domingos, Pedros. 2012. "A Few Useful Things to Know about Machine Learning". *Cummun acm* 55 (10): 77-87.

Levine, Sergey et al. 2018. "Learning hand-eye coordination for robotic grasping with deep learning large-scale data collection". *The International Journal of Robotics Research* 37 (4-5): 421-436.

Field, Matthew et al. 2009. "Motion capture in robotics review". *IEEE International Conference on Control and Automation*: 697-1702.

Miller, Nathan et al. 2004. "Motion capture for inertial sensing for untethered humanoid teleoperation". *IEEE-RAS International Conference on Humanoid* 4: 547-565.

Breazeal, Cynthia & Scassellati, Brian. 2002. "Robots that imitate humans". *TRENDS in Cognitive Sciences* 6 (11): 481-487.

Wang, Peng; Liu, Hongyi & Gao, Robert X. 2018. "Deep learning-based human motion recognition for predictive context-aware human-robot collaboration". *CIRP Annals - Manufacturing Technology*: 17-20.