2009

Motion capture in robotics review

Matthew Field  
*University of Wollongong, field@uow.edu.au*

David A. Stirling  
*University of Wollongong, stirling@uow.edu.au*

Fazel Naghdy  
*University of Wollongong, fazel@uow.edu.au*

Zengxi Pan  
*University of Wollongong, zengxi@uow.edu.au*


Publication Details  
Motion capture in robotics review

Matthew Field, David Stirling, Fazel Naghdy
School of Electrical, Computer and Telecommunications Engineering
University of Wollongong
NSW, Australia
Email: {mf91, stirling, fazel}@uow.edu.au

Zengxi Pan
School of Mechanical, Materials and Mechatronics Engineering
University of Wollongong
NSW, Australia
Email: zengxi@uow.edu.au

Abstract—This survey reviews motion capture technologies and the current challenges associated with their application in robotic systems. Various sensor systems used in current literature are introduced and evaluated based on the relative strengths and weaknesses. Some research problems pursued with these sensors in robotics are reviewed and application areas are discussed. Significant methodologies in analysing the sensor data are discussed and evaluated based on the perceived benefits and limitations. Finally, results from experimentation with an inertial motion capture system are shown based on clustering and segmentation techniques.

I. INTRODUCTION

Motion tracking is a vital component of developing intelligent autonomous robots. A robot agent must be able to perceive human motion in order to interact, co-operate, or imitate in an intelligent manner. In recent years tracking technology has become increasingly miniaturized, and along with computing power, more available. Measurement error is reduced to a minimum with advanced tracking algorithms and post-processing. Several different sensing mechanisms can track sufficient motion leading to wide variety of uses in autonomous systems. Essentially motion is recorded by tracking the precise position and orientation of points of interest at high frequency. Each tracker uses fundamentally different physical principles to measure position and orientation. Mechanisms vary from using a multiplexed reading of orthogonal magnetic fields from inductive coils, accelerometers and gyroscopes, intensity of ultrasonic pulses, the mechanical orientation of joints, or a reconstruction of the position of visible markers detected with multiple cameras. This results in systems with varying capabilities and susceptibilities, such as occluded trackers, constrained motion or magnetic disturbances.

For robotic and automation applications include surveillance and human interaction, which requires a form of identity or action recognition [39][5], teleoperation [23], robot programming by demonstration [8][2] or humanoid imitation [26].

Apart from sensing motion and collecting data, research in recent years has focused on methodologies in handling large high dimensional data sets which arise from multiple sensors. Ultimately the sensing leads to classification or to controlling an external device, through the use of a range of analysis techniques from machine learning.

The purpose of this review paper is to collate a compendium of recent approaches to human motion tracking in the context of robotic research in order to highlight potential advantages of each sensing mechanism. It is important to track progress in this area since the technology is new and changes quickly especially for current trends in analysis methodology. This paper describes various research and commercial tracking technology, their implementation and significance in robotic research and briefly describes some current analysis of large data sets produced by these pervasive technologies.

The paper is organised as follows. Firstly, a description of previous surveys and where this paper fits is discussed, section II outlines the various tracking technology available grouped by the sensing mechanism. Section III describes in further detail the applications in robotics that have risen out of these technologies. Section IV shows a comparison of the features of the sensors. Section V displays experimentation with an inertial motion capture system and section VI concludes the paper.

A. Previous surveys

There have been few surveys of motion tracking sensors in recent years. [40] gave a tutorial on sensing technology range and focused on augmented reality applications. The most comprehensive is [24] which is primarily focused on advances in image processing for markerless motion capture rather than considering wearable marker motion tracking. However, significant motion capture advances were also summarised including a number of analyses of motion capture data. There have been a number of other vision based surveys which devote minimal attention to wearable sensors [15].

Other reviews make sparse references to previous work in the robotics field. This review is intended to provide a targeted approach and specifically summarize recent approaches of motion tracking in robotics. Various analysis methodologies are compared in terms of their benefits and limitations.

II. TRACKING TECHNOLOGIES

Available tracking techniques have been categorised based upon the working principle. Although markerless motion analysis is a highly active research area, this paper focuses more on the use of wearable sensing in a robotics context.

Each motion tracking system has advantages that are useful depending on the application. Table I lists the major features
of each type of sensor system and references some researchers using the technology in robotics or automation.

A. Optical - Passive Marker

Optical detection with passive markers, or reflective indicators, uses multiple fixed high speed cameras around the measurement area to triangulate a precise marker position. Infrared lighting allows the capture of high contrast images of the reflective markers up to 2kHz. At least two cameras at a time must capture a marker otherwise there are occlusion errors. Although markers cannot be differentiated from each other until post-processing analysis restores the correct path. This results set of unlabeled points in a three dimensional workspace that correspond to the kinematic structure of the subject.

These optical systems are affected by instances of occluded markers but successful recordings have sub-millimeter errors. Redundant markers are often used to overcome occlusions which reduces the probably of error but increasing the number of markers also increases the processing latency. An advantage for passive marker systems is that the subject is not weighed down with battery packs or constrained by wires to sensors. Some significant disadvantages include portability and the measurement workspace, which is a small fixed area in view of the cameras. The area can be increased but this is still limited by the space in an indoor venue and strength of the reflected light.

B. Optical - Active Marker

Active optical markers act as a light source instead of a reflector and are often deployed as infrared emitting diodes (IREDs). The light emission from markers is multiplexed and therefore the frequency of the camera speed is divided by the numbers of sensors to detect. Although this introduces a limitation on measurement frequency, less post-processing required since individual LEDs can be identified.

Once again the capture is limited by the arrangement of cameras and the field of view. The measurement area is typically in the order of several square meters, and theoretically higher than for passive systems because of the light intensity diminishing with inverse square of distance. Since the indicators are powered, for wireless recording the subject must wear power packs and secure wires that would otherwise impede motion.

C. Optical - Markerless

Ideally motion capture would only use one set of camera(s) from one angle, similar to human vision, without requiring any body markers. Although these vision-based processing techniques are a topic of research the only accurate systems are confined to a restricted area and background, generally provide inaccurate estimates or require cameras from multiple viewing angles. Due to the extensive research in vision-based processing a more in depth survey targeting this research can be found in [24][15].

Markerless motion capture is an ongoing research area with massive potential. It relies upon image segmentation and processing techniques to find a human posture which may be matched to a human template [36]. Common approaches employ background scene subtraction techniques to extract a silhouette [22] and various manifold learning algorithms [12].

D. Inertial

Inertial motion capture relies on acceleration and rotational velocity measurements from triaxial accelerometers and gyroscopes. Each inertial sensor positioned at strategic points on the body measures precise orientation to within $2^\circ$ RMS [10]. This is achieved with estimation techniques such as Kalman filtering [31] fusing the angular rate with incline (gravity vector) and, for some sensors, magnetometers for more reliable heading data. Assuming certain configuration for the sensors and calibrating the actor dimensions an accurate posture can be resolved.

A major drawback of these sensors is estimating the position by integrating accelerations or angular velocity, a cumulative error arises, referred to as drift. Modern inertial motion capture suits rely upon ground contact force detection, indicated by sudden foot accelerations, to update reference position. Without well defined events such as these the posture remains accurate but tracking world position is unreliable due to drift. Other limitations include the need for post-processing in uncertain environments, when the ground support is varying dramatically.

Despite inherent problems associated with this technique it is improved in combination with other technology. [37] used inertial sensors with ultrasonic detection for a practical outdoor capture technique. With one optical marker the suit may be tracked accurately within the camera workspace.

E. Magnetic

Electromagnetic fields are established through precise current pulses in mounted transmitting antennae. Each magnetic
field including the earth magnetic field is measured giving an estimation of joint position, angles and global orientation. AC electromagnetic systems are highly distorted by neighbouring metallic objects but recent DC magnetic field systems exhibit significantly reduced distortion.

A triaxial transmitter produces DC pulses sequentially to each axis and the receiving antennae, mounted on significant positions on the body, measure the magnetic field along each axis. The earth magnetic field is measured when no pulse is present and subtracted when measuring the orientation. This results in 6DOF position and orientation information for each sensor up to a range of 10 ft from the transmitter.

Advantages of this approach include the flexibility in locating the sensors on the body, there are no occlusion issues. The measurement area is limited to a small region around the transmitter, comparable to optical systems, and is as portable as the transmitter. Metallic objects still cause a significant level noise and distortion to measurements.

F. Mechanical

The simplest method of capturing pose is to measure orientation directly using electromechanical potentiometers measuring the orientation displacement of each joint. This approach is effective in many cases it is not affected by external forces or occlusions, measurements can be fast and the equipment portable. The main disadvantage is that motion is usually constrained by the rigidity of the wearable equipment. An exo-skeletal frame normally imposes restrictions on the range of motion since human joints are more flexible than the mechanical links. Another problem is in detecting the true position and orientation of the entire frame. This mechanism cannot detect events such as jumping or turning, only the relative angle between limbs. Therefore captured results appear to slide, a problem that can be overcome by incorporating other measurement techniques. This method is particularly strong in exoskeletal frames and prosthetics since the joints must also be powered.

G. Acoustic

By attaching ultrasonic transmitters and microphones at specific locations on a moving body an estimate of position can be determined through the intensity of acoustic pulses. The pulses are multiplexed so that each microphone measures the pulse intensity from each transmitter to estimate the relative distances between all sensor points.

A complication arising from this arrangement is self-occlusion, that is, parts of the moving body blocking a direct path to receiving microphones. It is especially difficult with partial occlusions since the reduced intensity should not be related to distance. Depending on the frequencies used the system is susceptible to background noises, temperature and humidity in uncertain environments, and to wind when used outdoors.

III. APPLICATIONS

Some of the major research applications of motion capture in robotics include programming by demonstration, imitation, tele-operation, activity or context recognition and humanoid designs. In Table II, significant methods used in analysing the motion data are compared.

A. Programming by Demonstration / Imitation

Robot programming by demonstration has a relatively long history. Research into faster programming of industrial robots has extended to imitative robotics in recent years by using motion capture technology and machine learning. Initially demonstrated trajectories could be followed by extracting key points or way points for the end effector allowing a demonstrator to show a particular path and the robot to follow by targeting the key points [25]. This resulted in a brittle control scheme where a robot could replicate motions but would fail in a different context.

Further research advances from following trajectories to learning and generalisation of motor manipulation skills, to imitate humans in a flexible manner. Motion is still often assumed to be composed of an arrangement of more primitive components, or motion primitives. A range of stochastic models and sequencing algorithms are typically used to learn motion primitive and generate suitable trajectories.

A common stochastic model for analysing human motion is the Hidden Markov Model (HMM) [18]. A number of HMM states were trained on motion capture sequences such that each state embodied a posture for the robot. States were compared in a ‘proto-symbol’ space and merged based on their relative Kullback-Leibler distances. [21] expanded upon this framework by incrementally updating the model and creating a hierarchy of HMM sequences using Factorial HMM (FHMM).

Another method is to transform high dimensional data into a low dimensional manifold using analyses such as Principal Component Analysis (PCA) or non-linear methods such as Isomap and Gaussian Process Latent Variable Model (GPLVM) which have shown higher performance in capturing relevant data structure. Non-linear dimension reduction techniques as used by [38] embed the data onto meaningful planes of motion style and content with relatively small data sets. Their methods based on GPLVM could sample from regions of the latent space where there were no observed data. This algorithm has also been implemented in humanoid imitation [33] by projecting data from the latent space on the robots reduced DOF.

[8] used an arrangement of inertial motion trackers on the upper body to capture arm and torso motion. Over many demonstrations the data was compressed by a PCA pre-processing step and clustered similar postures into a GMM of a size determined using the Bayesian Information Criterion (BIC). Generalized trajectories could be restored and reproduced in a humanoid in different contexts by using Gaussian Mixture Regression (GMR) between the appropriate sequence of states.

There has been full body humanoid imitation [30] captured human motions with an optical passive marker system and translated the angles into a frame to replay motion. [26] used human motion capture of a traditional dance to control a


<table>
<thead>
<tr>
<th>Method</th>
<th>Advantages</th>
<th>Disadvantages</th>
<th>References</th>
</tr>
</thead>
</table>
| Optical - Passive | - Precision < 1mm  
|                 | - Wireless  
|                 | - Less burden  
|                 | -                                      | - Position only  
|                 | - Limited measurement space  
|                 | - Occlusions  
|                 | - Post-processing latency                  | [11][33][18] |
| Optical - Active | - Precision < 1mm  
|                 | - Wireless  
|                 | - Higher range than passive  
|                 | -                                      | - Position only  
|                 | - Limited measurement space  
|                 | - Occlusions  
|                 | - Post-processing latency                  | [21][12] |
| Optical - Markerless | - Wireless-Outdoor  
|                 | - Flexible  
|                 | - No sensor burden  
|                 | - Contextual information  
|                 | -                                      | - High noise  
|                 | - Occlusions  
|                 | - High post-processing cost  
|                 | - Generally not real-time  
|                 | - High sensitivity to lighting               | [3] |
| Inertial | - Accelerations  
|                 | - Precision < a degree  
|                 | - Wireless - outdoor  
|                 | - Fast calibration  
|                 | - Portable  
|                 | -                                      | - No reference position  
|                 | - Post-processing - external contacts  
|                 | - Noise  
|                 | - Magnetic disturbances                      | [8][13][23] |
| Mechanical | - Portable  
|                 | - Wireless-outdoor  
|                 | - Robust, reliable  
|                 | -                                      | - Restrictive movement  
|                 | - No reference position  
|                 | - Relative orientation only                 | [17] |
| Magnetic | - Portable  
|                 | - Wireless-outdoor  
|                 | - Flexible sensor arrangement  
|                 | -                                      | - Limited range  
|                 | - No reference position  
|                 | - Magnetic disturbances                      | [28] |
| Acoustic | - Portable  
|                 | - Wireless-outdoor  
|                 | - Flexible sensor arrangement  
|                 | -                                      | - Partial occlusions  
|                 | - No reference position  
|                 | - Environmental conditions                  | [37] |

[27] transferred modified human motion capture data into humanoid simulations. [7] and [1] provide a good reviews of robotic imitation approaches.

Direct real-time mapping of human motion to robots has many applications in teleoperation tasks. Miller [23] used a set of inertial sensors to control the robot arm of NASA Robonaut. Only 3 sensors were used for untethered control.

### B. Activity Recognition

Understanding observations is an important aspect of autonomous systems and a significant amount of research in recent years has been devoted to identifying people and classifying their actions, as evident in surveys [24]. The action recognition problem has been pursued by researchers from many disciplines due to significant potential applications but research especially with video sequences is still in its infancy [22].

[39] uses a few inertial sensors and microphones placed on one arm to identify activities within a greater task. This is for the purposes of assistant computing which recognises, given the context of the measurements, what task is being performed. It can therefore provide relevant information, for instance, an assembly manual in a manufacturing or workshop environment as used in the paper. [41] used similar approaches in automotive repair environment.

Behaviour segmentation is also a recurring theme in computer animation research. [5] mined motifs in large motion capture databases by clustering posture using k-means to create structured graphs which can blend fluid animations. Segmentation techniques were also evaluated by [4] for automating motion capture editing, the most successful approach involved Probabilistic PCA (PPCA) and Mahalonobis distance thresholding to segment plausible actions.

### C. Humanoid Design

Motion capture measurements are an important resource for humanoid interactions and learning. Observations of human actions are a major influence in humanoid robot designs. This ranges from informing stable bipedal designs, energy conservation of actions to learning control actions in stabilization.

Motion capture has led to translating human gaits into humanoid motion by adapting the trajectory guided by the ZMP constraint [11] or by compensating for the angular velocity of the pelvis to stabilise the frame [27].

Other research in constructing stable, efficient robotic frames have been influenced by biomechanics research which in turn uses motion capture. [9] constructed mechanical biped frames that could walk passively down a small incline. This demonstration led to increased study of the mechanical design of walking bipeds [16] including conclusions from biomechanics.

### IV. Experimental Results with Motion capture

At the University of Wollongong we have experimented with an inertial motion capture system from Xsens Technolo-
TABLE II

<table>
<thead>
<tr>
<th>Methodology</th>
<th>Benefits</th>
<th>Limitations</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Key framing</td>
<td>Simplicity</td>
<td>No model of process</td>
<td>[25]</td>
</tr>
<tr>
<td></td>
<td>- Low memory consumption</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clustering - GMM</td>
<td>Probabilistic model</td>
<td>Model density uncertain in high dimension thereby difficult to train</td>
<td>[8][13]</td>
</tr>
<tr>
<td></td>
<td>- Incremental model</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Fewer parameters</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Symbolic</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- No model of process</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Cannot sample unobserved space</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clustering - HMM</td>
<td>Probabilistic model</td>
<td>Poor trajectory generation</td>
<td>[18][21][39]</td>
</tr>
<tr>
<td></td>
<td>- Implicit dynamics</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Incremental model</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Symbolic</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Fewer model of process</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Cannot sample unobserved space</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Explicit dynamics</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clustering - K-means</td>
<td>Simple Euclidean separation</td>
<td>No model of process</td>
<td>[5]</td>
</tr>
<tr>
<td></td>
<td>- Fast processing</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Low memory consumption</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCA</td>
<td>Fast processing</td>
<td>Reliant on variance</td>
<td>[8]</td>
</tr>
<tr>
<td></td>
<td>- Simple to implement</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Used as pre-process</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Not probabilistic i.e. cannot sample</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-linear dimension reduction (Isomap, GPLVM etc.)</td>
<td>Non-linear mapping</td>
<td>Difficult to compare latent model</td>
<td>[19][33][38]</td>
</tr>
<tr>
<td></td>
<td>- Probabilistic model</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Generalize with minimal data</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Linear mapping to latent variables</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Connectionist</td>
<td>Biological premise</td>
<td>Require large data sets</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Prediction performance</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Model</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Require large data sets</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Model</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In [14] it was shown that removing less abundant clusters from the GMM hinder identification of the activity and is detrimental to a stable center of mass trajectory in possible robot motions. In [13] recognisable behaviour was segmented with a range of techniques and compared to subjective segmentations. A close relationship was shown between the algorithm segmentation and human judgement. These techniques may work towards recognising particular activities, discovering anomalous behaviour or assisting humanoid imitation planning.

V. CONCLUSION

Recent applications of motion tracking technology in robotics have been presented. The various advantages and disadvantages of each sensor mechanism are compared revealing inherent limitations of the technology. Every sensor has drawbacks but some combined sensor applications appear to overcome these problems to some extent. The relevance of motion capture for robotics was discussed and some current techniques in data analysis were outlined to illustrate the difficulty in handling this data rich sensor technology. The trend in methodology is towards stochastic machine learning techniques such as HMM or GMM and non-linear dimension reduction. The resulting empirical models tend to handle uncertainty well and are suitable for incrementally updating models. Finally, some recent experimental work using an inertial motion capture system is outlined along with the methodology employed to analyse the data.

Markerless motion capture is undeniably important for future robotics and automation research, however robot learning even with accurate motion capture is limited. Among the challenges in human-robot interaction today include expanding upon generalising motions to understanding motion planning and decisions and building ultimately context aware systems. The technology outlined in this survey provides sufficient data to approach the problem. Methodologies in handling the data are generally limited in their scope and application.

REFERENCES
