Human Motion Capturing

Autors	Herzog Michael, Kratzer Fabian & Mayer Anjela
Betreuer	Isabel Ehrenberger
Bearbeitungsdauer	ca. 90 Stunden
Präsentationstermin	08.07.2018

Introduction/Motivation

Motion in robotics is a difficult task, concerning the complexity of robot kinematics, dynamics and their environment. When it comes to robot motion, it is necessary to plan the desired movement beforehand. It is required to generate a trajectory of the motion, which contains sequential positions of the robot over time. Afterwards the positions are mapped onto the joint angles of the robot [3].

Additionally, for human robots, the motion should appear natural for human observer. Irritating movements are to avoid, stating even more difficulties to generate movement trajectories [1]. Therefore, the most obvious approach for humanoid robots is to capture human motion and map it to the robot by using appropriate models, also referred to as motion capture. Motion capture, also known as motion tracking, requires position information over time of the designated subject. There are several different methods to implement the position tracking.

During the seminar "Motion in Man and Machine" in 2018 at the KIT, we, as a group of three students, wanted to become acquainted with the topic "Human Motion Capturing". Therefore, we did research to understand the theoretical fundamentals of motion capturing and to determine the state of the art. In the hands-on part of the seminar we did our own motion capture in the Vicon Lab with the master motor map framework (MMM) [2] of the H²T Institute.

In the following our work is presented, starting with an overview of the different methods of motion capture. Afterwards, the way from capturing motion to the mapping to the MMM is described. The text provides the theoretical basis and explanations of motion tracking. The podcast video illustrates the consecutive steps. The single steps of the podcast video are further described in the text.

Podcast

m3video_small1.mp4

Theoretical Basics

In this section, the basics of motion tracking and data processing are described. The different methods of tracking and the data processing steps to perform are described. The last part of the theoretical basics is an introduction to the MMM framework.

Motion Capture Methods

The following part gives an overview on the different methods for motion capture. They can be distinguished by optical and non-optical methods. Optical methods use image-based systems to determine the spacial positions. They are subdivided into marker-based motion capturing, with active or passive markers, and marker-less motion capture. Motion capture methods applying inertial, magnetic, mechanical and acoustic sensors are grouped in non-optical motion capture.

Optical Motion Capture

Optical methods utilize several cameras with different perspectives on the subject. The spatial position is determined by recognizing the subject features and triangulating their positions within the image from different perspectives. Accordingly, the complete motion is derived from the captured sequence of images. For example, figure 1 illustrates two projections, X1 and X2, of the position X on the image panes of two different cameras. With the given coordinates of X1 and X2 the original position of X can be determined with the help of epipolar geometry. Optical motion capture methods can be classified furthermore into marker-based and marker-less tracking.
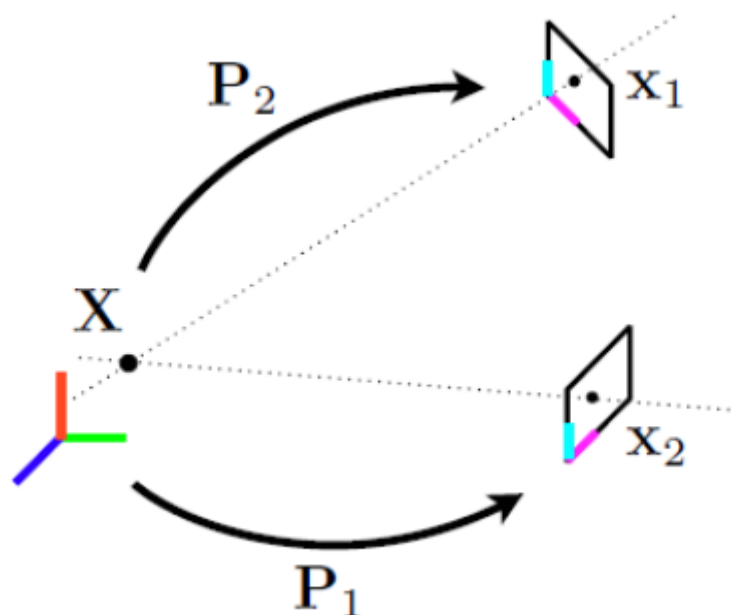

Figure 1: Position X mapped to the image planes of two different perspectives [4]

Marker-based motion capture

Marker-based methods for motion capture require active or passive markers to be attached to the subject performing the movement. While passive markers are recognized by reflecting the light which is generated near the cameras, active marker are emitting light on their own. To ensure that passive markers are visible to each camera it has to be illuminated from the direction of each camera, hence passive Systems have lights mounted on the cameras. Passive markers are coated with a retroreflective material which wears off over time so that the markers or their coating have to be replaced. Since all markers in passive systems looks similar the effect of marker-swapping is more likely to occur than with active systems. The LED's of an active system, on the other hand, can be activated one at a time which helps to identify the markers. On contrary to passive markers the

Intensity of the light emitting from active markers is more powerful which increases the volume and distance for capture. It also enables capturing in environments with difficult lighting and is suitable for outside capture. As a drawback active marker require some wiring for the LED's and are less handy to wear. As a conclusion, systems using active markers are significantly better at finding correspondences while systems working with passive marker require more manual cleanup after the capturing [4].

Marker-less motion capture

Marker-less motion capture relies on computer vision algorithms, so that no additional applications on the moving subject are needed. The motion is recorded by multiple video streams and analyzed by algorithms in order to recognize human shapes and track the movements of their body parts [5]. Because no markers are used, this method is non-intrusive while it should be highly accurate since every pixel on the body is captured [4]. However, marker-less motion capture is still in research phase and its accuracy is highly dependent on computer vision algorithms. In order to enhance computational performance, simplistic or generic human models are used, which results in inaccurate joint information and finally inaccurate movement tracking [13]. Furthermore, human body self-occlusion is a major cause of ambiguities in body part tracking [13].

The main advantage of optical systems in general is the very high sampling rate, which enables the capturing of fast movements [5]. In the marker-based motion capture markers are often not recognized, for example when being hidden by other objects or by the subject itself. In this case the missing position information needs to be reconstructed manually with the help of algorithms in a time-consuming process. Marker-less motion capture avoids difficulties with physical markers completely but faces different challenges on the software side.

Non-optical Motion Capture

Besides optical methods for motion capture, there are also several non-optical technologies which can be differed by the applied sensors.

Inertial Motion Capture

Inertial motion tracking works with a set of Inertial Measurement Units (IMUs) and biomechanical models. The IMUs measure rotational rates during the motion and therefore mostly consist of a gyroscope, magnetometer and accelerometer. For motion capture there are, depending on the intended application, different amounts of IMUs distributed over the body and wired together into a master-slave system. The master synchronizes all sensor sampling, provides them with power and handles the communication with the PC [6]. There are wireless as well as wired systems available. With the help of sensor fusion algorithms, the sensor data can then be merged and mapped onto a joint model. This method of motion capturing does not require cameras or marker and is inexpensive, but a drift of the position can occur since no global position can be measured with IMUs [4].

Magnetic Motion Capture

Magnetic motion capture utilizes external transmitters to establish magnetic fields in space, which

can be measured by sensors placed on the subject. The sensors and transmitter are wired to a control unit which correlates their measured locations within the field and transmits the data to a computer [10] wirelessly or across wire [4]. The joint angles of the subject are determined with the help of inverse kinematic. Magnetic motion capture can handle hidden subjects well and provides a low-cost system compared with the optical motion capture system. On the other hand, the range of the system is highly restricted and interferences from metallic objects can occur [11].

Mechanical Motion Capture

Mechanical motion capture systems consist of potentiometers and sliders put together into a suit. The sensors can estimate joint angles directly when positioned close to the joints on the subject's body [12]. Potentiometers record analog voltage changes and convert them to digital values. Because of that they are only able to measure changes from the original orientation, so no global position is determined and the calibration is difficult [4]. Additionally, the sliders and potentiometers restrict the subject's freedom of movement. Advantages of mechanical motion capture systems are their low cost and high flexibility, since they are self-contained. Finally, they are also free-of-occlusion, mechanical motion capture is not affected by magnetic fields or unwanted reflections [5].

Acoustic Motion Capture

In an acoustic motion capture system, a set of sound emitters is placed on the subject which are sequentially activated producing characteristic frequencies. On the capture site the receptors are on fixed positions receiving the frequencies and calculating the positions of each emitter. The location of the emitters is determined by its distance to different receiver by using the time-of-flight of the acoustic signal in the calculation. Acoustic motion capture systems have high accuracy but are prone to signal interferences [15]. Moreover, the location of the receivers has to be calibrated and the wiring [15] of the emitter restricts the degree of freedom [5].

Unlike optical systems, the non-optical systems do not require a complex camera and lighting setup to capture motions. For this purpose, different sensor data is used to determine positions. Each capture method has its own difficulties, advantages and disadvantages, so it depends on the application which one of these methods is the most suitable. Finally, to add captures resulting from the different capture methods to the MMM Framework, it needs particular converters for each method which are described in the following.

MMM Framework

The MMM developed at the H²T Institute at KIT [2] is a framework which contains a generalized reference model of the human body and provides tools to handle sampled data and to make them available to output modules. Sampled data such as data from motion tracking are converted to a human reference model of one meter height. The reference model uses the markers at the same relative position of the human body as the human that is captured. That means that regardless who is converted, every conversion ends up with the reference model. The model itself is based on statistical data of the human body. The individual properties of the person captured, such as leg length or weight, are converted to match the reference model. Data that are converted to the MMM can also be saved in the Motion Database (DB) for later use. The MMM representation of the motion captured can

be converted to every arbitrary robot. For that, a converter tool using the provided interfaced is needed. After the mapping, maybe there is need to slightly adapt some parameters, i.e. for the engines, because the technical properties of the robot model can differ a little bit to the one of the real robot. Figure 2 illustrates the workflow explained above and the various interfaces. As illustrated with the arrows pointing from the captures to MMM, as soon as the captured data is converted, it can be further progressed by all converters illustrated on the bottom or it can be saved to the Motion DB and thus used later. This approach is very flexible as several methods to obtain samples and to post-process the model calculated can be used. Furthermore, it can be extended and thus adapted to new techniques or requirements.
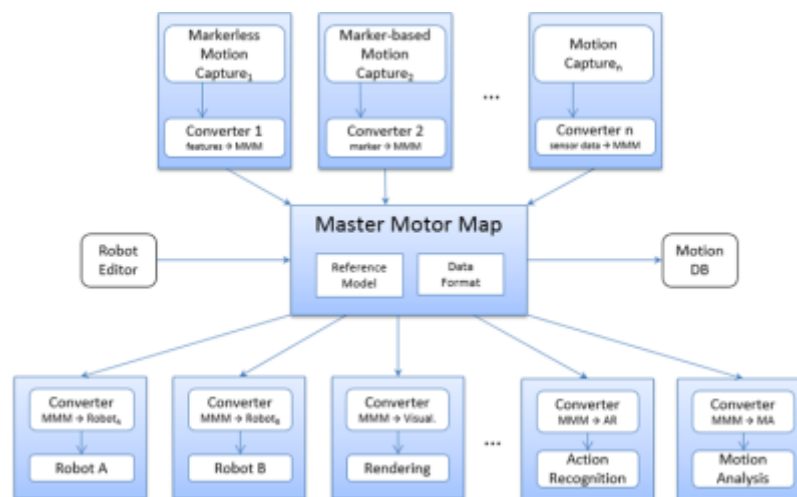

Figure 2: The MMM framework [1]

Experimental Setup

The VICON lab was used to gain marker-based motion capture data. In this lab, motion capture data was recorded and post-processed in order to make it transformable to the MMM. Four setups were performed. Firstly, a rigid object (a packaged sponge) was moved in space. Afterwards, a bottle as a second object was added. They were then again moved in space, also crossing each other's ways. At last, a human wearing a suit with markers, first picked up the sponge and then threw it away. The bottle was no longer used in this scenario. This methodologic approach was used in order to increase the complexity of the recording and data processing step-by-step.

Cameras and calibration

The subjects were recorded using a VICON NEXUS camera systems with MX10 cameras facing the scene from approximately two meters above. Thus, it was an optical tracking with passive markers. Each recording was performed three times to make sure at least one recording allowed further processing. At the same time, a video camera was used to record the scene. This makes an interpretation of the obtained results easier.

Before recording data, the cameras must be calibrated. So that the recording system knows how the cameras are placed to each other and where the ground begins. Therefore, all cameras focus on the calibration bar, which figure 3 shows. Also, contrast and light intensity threshold can be adjusted. This makes sure that the system uses only the desired markers and no artefacts.

Figure 3: The calibration stick

All objects have a fixed marker set. This means the markers on the objects always have the same relative position on the object. Many objects are stored in a database along with the positions of their markers. Later, every time an object appears in the recording scene, the system will search the database for an object which marker positions are the same as those of the object in the scene.

Humans wear a suit with markers placed on. Those markers are placed on well-defined landmarks of the human body which figure 4 shows. In total, there are 51 markers which positions are further described in [14]. The markers' positions are also stored in the VICON software database.

In a perfect recorded scenario, all the markers on the human match those of the software.


Figure 4: Man wearing the suit with markers [14]

Data preparation pipeline

Usually, the recording software does not correctly label all the found markers. This happens for example if markers are hidden by other body parts or if they are in the shadow so that they cannot reflect the light. In those cases, the following steps are necessary to get a complete and usable representation of the recorded data. The steps are explained according to the VICON documentation [8].

First, the "Core Processing" step aims at producing 3D trajectories with automatically reconstruction and kinematic fitting. Afterwards, "labeling" tries to label all markers by comparing all detected markers to the markers of the stored subject models. Usually, not all markers can automatically be labeled. Therefore, the left markers have to be labelled by hand. "Fill Gaps" tries to reconstruct the trajectories as good as possible with the new manually labeled markers. This is done by an interpolation over known positions. Sometimes, makers are displayed which are not genuine markers, but artefacts. Those can be removed with the "Delete Unlabeled Trajectories" command. Afterwards, two filtering methods (Woltring and Butterworth) are applied. Woltering uses a quintic spline interpolation and Butterworth a low-pass filter to filter out signal noise above 300 Hz. They aim at obtaining smooth trajectories. "Fit subject motion" fits the recorded motion as a whole instead of frame-by-frame as in the Core Processor. Finally, "Export C3D" saves data to a .c3d file [7].

Conversion to the MMM

The last step, we performed in the lab was the conversion of the completely labeled captured motion to the MMM model. We used a converter that comes with the framework. This converter solves a minimization problem to match as good as possible the recorded marker positions and the virtual markers of the reference model [9]. To give the converter some information at the beginning, the recorded person's height and weight are provided. This simplifies the solving of the minimization problem [1]. The resulting representation of the human by the MMM framework is shown in figure 5. On the left side, the recorded human with the sponge in his hand is completely labeled and on the right side, it is shown after the conversion process. As explained in the previous section, this MMM representation of the captured motion can now be converted to any desired robot.
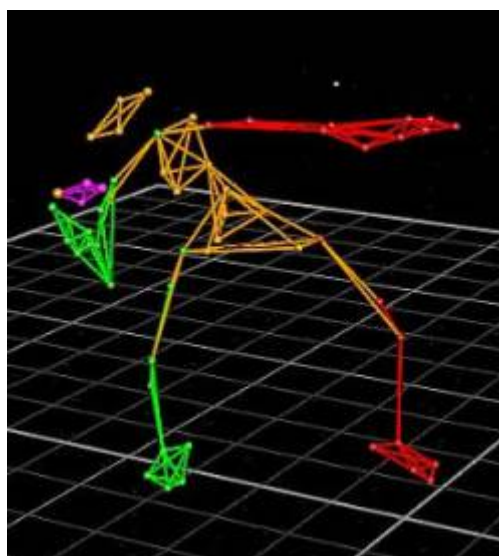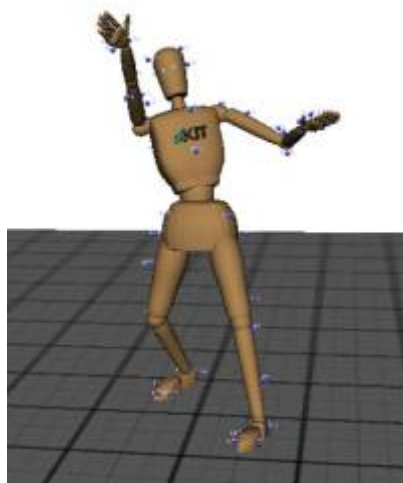
Figure 5: Top: the recorded human with the sponge in his hand. Bottom: the movement is mapped to the MMM model.

Difficulties

In this chapter, the problems occurred in our lab session are described. The biggest problem was to fill trajectories when the markers were not detected in every frame. This happened when we recorded a movement where the markers faced the floor and were hidden by other parts of the human body. As we used only cameras that were installed in two meters height, those markers could not be recorded. In particular, this was a problem when we tried to fill the trajectories of the markers placed on the hand. There are many markers situated close to each other. Thus, there were consecutive frames where it lacked up to five markers. In that case, we could not reconstruct the trajectories properly.

Another problem was revealed not before the final mapping to the MMM model. Some markers are supposed to be placed on both body sites on the same height. If there was a difference, the model would look skewed. Then the movement of the model was strange, though the human's movement was not.

We learned that a good set-up of the markers and cameras helps save time for the post-processing steps. The markers should as good as possible be recordable by the camera. Here, experience in recording can help improve the quality of the recordings.

Summary and Conclusion

In our lab session we performed the whole process of motion tracking. First, we selected scenarios, then we set-up the cameras and the environment. The next step was to carefully calibrate the cameras. Then, we recorded three scenarios three times each. Afterwards, we post-processed the recordings with the help of the VICON software. Here, we spent some time fixing trajectories as described above. Finally, we converted the post-processed recording to the MMM model. The steps to perform from recording a movement to the ready-to-use MMM model are numerous and took us several hours. Even though we took care of proper positioning of the markers and that the movement

could be fully recorded, at the end it turned out that some markers were hidden or placed incorrectly. In order to obtain high quality recordings, experience is helpful to avoid those problems. Additionally, with cameras placed on the floor, the quality of our recordings would have been better.

References

[1] Master Motor Map: MMMCore Documentation. (2018, May 6). From:
https://mmm.humanoids.kit.edu/index.html

[2] Azad, P., Asfour, T., & Dillmann, R. (2007, April). Toward an unified representation for imitation of human motion on humanoids. In Robotics and Automation, 2007 IEEE International Conference on (pp. 2558-2563). IEEE.

[3] Yamane, K., & Hodgins, J. (2009, October). Simultaneous tracking and balancing of humanoid robots for imitating human motion capture data. In Intelligent Robots and Systems, 2009. IROS 2009. IEEE/RSJ International Conference on (pp. 2510-2517). IEEE.

[4] Sigal, Leonid, Carnegie Mellon University. (2012). Human Motion Modeling and Analysis: Lecture 3 (Marker-based) Motion Capture [Lecture]. (2018, May 6). From:
http://www.cs.cmu.edu/~yaser/Lecture-3-MarkerBasedMocap.pdf

[5] Ashish Sharma, M., & Ashish Sharma, P. (2013). MOTION CAPTURE PROCESS, TECHNIQUES AND APPLICATIONS. International Journal on Recent and Innovation Trends in Computing and Communication, 1(4), 251-257. Abgerufen von www.ijritcc.org/download/IJRITCC_1350.pdf

[6] Roetenberg, D., Luinge, H., & Slycke, P. (2009). Xsens MVN: full 6DOF human motion tracking using miniature inertial sensors. Xsens Motion Technologies BV, Tech. Rep, 1.

[7] VICON. Vicon Nexus Product Guide.(2018,June 7). From:
http://documentation.vicon.com/nexus/v2.2/Nexus1_8Guide.pdf

[8] VICON. Pipeline Tools. (2018, June 7). From:
https://docs.vicon.com/display/Nexus25/Pipeline+tools#Pipelinetools-FillGapOpsFillGap&FilterDataoperations

[9] Terlemez, Ömer, et al. "Master Motor Map (MMM)—Framework and toolkit for capturing, representing, and reproducing human motion on humanoid robots." Humanoid Robots (Humanoids), 2014 14th IEEE-RAS International Conference on. IEEE, 2014.

[10] Magnetic Motion Capture Systems. (n.d.). From:
http://metamotion.com/motion-capture/magnetic-motion-capture-1.htm

[11] Michael Gleicher. 1999. Animation from observation: Motion capture and motion editing. SIGGRAPH Comput. Graph. 33, 4 (November 1999), 51-54.
DOI=http://dx.doi.org/10.1145/345370.345409

[12] Motion Capture Sensor Systems. (2012, August 10). From:
http://www.azosensors.com/article.aspx?ArticleID=43

[13] Mündermann, L., Corazza, S., & Andriacchi, T. P. (2006). The evolution of methods for the capture of human movement leading to markerless motion capture for biomechanical applications. Journal of

NeuroEngineering and Rehabilitation, 3, 6. http://doi.org/10.1186/1743-0003-3-6

[14] H²T. Definition of the Marker Set. Retrieved from https://motion-database.humanoids.kit.edu/marker_set/ on 02 July 2018.

[15] Y. Zheng, K. C. Chan and C. C. L. Wang, "Pedalvatar: An IMU-based real-time body motion capture system using foot rooted kinematic model," 2014 IEEE/RSJ International Conference on Intelligent Robots and Systems, Chicago, IL, 2014, pp. 4130-4135.