

Czech Technical University in Prague
Faculty of Electrical Engineering
Department of Computer Graphics and Interaction



Bachelor Thesis

Humanoid Motion Capture

Kristina Sukhomlina

Thesis advisor: Ing. David Sedláček

Study program: Software technologies and management, Bachelor

Branch of study: Web and Multimedia

23. May 2014

Acknowledgements

I would like to thank my parents for their support, my thesis advisor Ing. David Sedláček for his great help, KME student Ekaterina Koshkina for her assistance with the testing, Ing. Zděnek Travníček for provided MoCap suit.

Prohlášení

Prohlašuji, že jsem práci vypracovala samostatně a použila jsem pouze podklady uvedené v příloženém seznamu.

Nemám závažný důvod proti užití tohoto školního díla ve smyslu §60 Zákona č. 121/2000 Sb., o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů (autorský zákon).

V Praze dne 23.5.2014

.....

Abstract

The following BA thesis explores humanoid motion tracking techniques, describes the steps required to collect and apply data on a basic 3D model and compares the outcomes. First goal is to create and prepare the humanoid model for the purpose of animation with the usage of Optitrack cameras and software provided by the Faculty of Electrical Engineering. Second task is to acquire the short takes for further processing through markerless tracking software, to evaluate, compare and contrast the results produced by both methods. The thesis also studies the influence of cameras' positioning on the quality of gathered data.

Abstrakt

Tato bakalářská práce se zabývá technikou sledování pohybu humanoida, popisuje řetězec zpracování dat od pořízení přes čištění až po následnou aplikaci na 3D model. Prvním cílem je vytvořit a připravit model humanoida pro záznam pohybu pomocí zařízení které se nachází v laboratoři VRlab (OptiTrack). Druhým úkolem je vygenerovat pět deseti sekundových sekvencí vhodných pro porovnání záznamu pohybu využívajících metod s a bez markerů, zhodnotit a porovnat výsledky. V práci se také analyzuje vliv umístění kamer na kvalitu získaných dat.

Table of Contents

Part I	
Introduction	10
Part II	
The Task	12
2.1 Modeling a humanoid	12
2.2 VRlab hardware description	12
2.2.1 Cameras	
2.3 Describing the data processing pipeline	13
2.4 Compare and contrast the outcomes	14
Part III	
The Research	16
3.1 Human body and its 3D graphical representation	16
3.2 Motion capture	16
3.2.1 Technology	17
3.2.1.1 Optical MoCap	
3.2.1.2 Non-optical MoCap	
3.2.2 Types of motion tracking systems	19
3.2.2.1 Passive optical system	
3.2.2.2 Active optical system	
3.2.2.3 Optical markerless motion capture	
3.2.2.4 Magnetic system	
3.2.2.5 Mechanical system	
3.2.2.6 Inertial system	
3.2.3 Data formats	20
3.2.3.1 C3D	
3.2.3.2 Comma-Separated Values or CSV	
3.2.3.3 BVH	
3.2.3.4 FBX	
3.3 Rigging	22
Part IV	
The Analysis	23
4.1 The software	23
4.2 Model	23
4.3 Testing the technology	24
4.4 Preparations with the actor	24
4.5 Applying the data on the rig	25
4.5.1 Motion Capture skeleton and Maya	
4.5 Markerless shooting	25
4.6 Cameras' positioning	26
Part V	

<u>The Workflow</u>	27
<u>5.1 Stage 0: Modeling and Rigging</u>	27
<u>5.1.1 Creating the rig</u>	27
<u>5.1.2 Placing controls (prev. model version)</u>	27
<u>5.2 Stages 1 and 2: Recording and Editing</u>	28
<u>5.2.1 Session 1</u>	28
<u>5.2.1.1 Editing</u>	
<u>5.2.1.2 The outcomes of Session 1</u>	
<u>5.2.2 Session 2</u>	30
<u>5.2.2.1 Calibrating</u>	
<u>5.2.2.2 The outcomes of Session 2</u>	
<u>5.2.3 Session 3</u>	31
<u>5.3 Stage 3: Application</u>	33
<u>5.4 Stage 4: Rendering</u>	34
<u>Part VI</u>	
<u>The Conclusion</u>	35
<u>Pictures</u>	39
<u>Initial Camera setup:</u>	
<u>Session 2, alternative camera setup:</u>	
<u>Joints in human skeleton:</u>	
<u>Rigging 'Smiley' version 1:</u>	
<u>Skin weights painting example:</u>	
<u>Camera calibration using the wand:</u>	
<u>OptiTrack Calibration Square:</u>	
<u>Motive:Tracker Calibration Layout example:</u>	
<u>Editing Phase, Session 2:</u>	
<u>Markerless MoCap processing - masking:</u>	
<u>Markerless MoCap - tracking in Kostilam:</u>	
<u>Rigging 'Smiley' version 2:</u>	
<u>MotionBuilder data Application:</u>	
<u>Disc Content:</u>	55

Part I

Introduction

Modeling humans or any kind of anthropomorphic creatures to be as naturally looking and moving as possible is a challenge for 3D artists and animators. Different studios and labs have made investments into development so that the accuracy has enormously increased. One of the approaches to make the 3D model's acting precise is Motion capture. The technology of tracking down the movements of real people is popular and widely spread. Although motion capture started merely as a tool for scientific purposes, it conquered several fields and even changed the perspective on CGI involvement in filmmaking.

The aim of this thesis was to study the process of human motion capturing using the equipment and hardware from VRlab of Czech Technical University, to test the provided technology's effectiveness, advantages and drawbacks, to apply and evaluate the gathered data. The 3D model and the control rig was created with respect to both manual animation and the one that was generated with the assistance of an actor. The data acquired using the Optitrack cameras from VRlab of Computer Graphics and Interaction department of CTU FEL was fully processed and exported so that it can be used for application on almost any kind of humanoid type of model. According to the task, five video sequences were shot and prepared for further markerless motion tracking development.

Part II

The Task

2.1 Modeling a humanoid

Humanoid model resembles human being in a sense of having a specific structure, which can be simply described by a tree graph. Both MotionBuilder and Maya software from Autodesk which were used for modeling and animation require this concrete hierarchy. The motion data from markers is mapped to the predefined joints of the model. These joints form dependencies that restrict the 3D model to certain motion rules, forming the skeleton of a humanoid type in a sense of specifying arms, legs, head and body, all starting from a root node.

Primitive model and its rig were created with Maya 2012 software under the Student license. The model ‘Smiley’, made for the study purposes of A7B39PRO, was changed to satisfy the MotionBuilder skeleton requirements. The model folder includes older version with manually developed Control rig and newer version used during Application stage.

2.2 VRlab hardware description

The VRlab of Czech Technical University provided with passive optical system, consisting of retroreflective markers, cameras, cables and essential software for motion capture experiments, whereas IIM lent the MoCap suit. Types of motion tracking systems are described in the text further.

The idea was to explore the newly installed tracking system that would be used for motion capturing with and without markers, to experiment with the cameras, Motive software from Optitrack and additional programs for markerless data processing provided by Ing. David Sedláček. The reasons behind the sessions in VRlab were to practically gain knowledge on motion capturing and also study the recently purchased hardware and software’s effectiveness.

2.2.1 Cameras

Motion capture cameras by OptiTrack were used to collect the data. So far VRlab possesses six s250e cameras with speed limit of 250 frames per second and Ethernet connectivity. Each camera covers 56 degrees angle and had its initial position at the ceiling. Group of six cameras shaped circle-like capture area. ([Initial Camera setup:](#))

As a part of the task the cameras were moved from its original position and recalibrated in order to study how their positioning would have influenced the quality of the takes. The actual outcomes are described further.

2.3 Describing the data processing pipeline

Transferring the real motion from the actor to a model is a complex process. The thesis covers the stages that each sequence of movements passes from footage to application. Stages were thoroughly described and their I/O can be found on the disc. It was decided that the motion sequences would be simple, involving just one participant. They do not include no interaction with side objects. This basic way enabled the clearer representation of the movement during the editing stage and also allowed to follow the differences in takes done with both original and alternative camera setup.

The motion sequences have stayed the same for each session with the exception of those marked as ‘improvisation’. Although, there was a constraint in space that could be used, the sequences came out various and distinctive from one another.

The editing and cleaning of data was made using Motive software by Optitrack and exported as .c3d files which were then imported into Autodesk MotionBuilder for animation.

Final stage was made using Autodesk Maya 2012. The takes from marker sessions were rendered as short videos with no additional montage, filtering or creative features since the emphasis was put on studying the technology and its functionality rather than artistic quality.

To avoid missing the details of each phase, the workflow description was divided into following sections:

0. Modeling and Rigging

1. Recording motion

2. Editing

3. Application

2.4 Compare and contrast the outcomes

It was stated that five takes should be made using the technique with and without markers, ten seconds each. However, during the markerless video recording the length of each take was extended due to one of the requirements of the processing software and as a safety precaution. The results are quite comparable, because, as already mentioned, the motion scenarios have been the same, although, the initial actor had to be replaced. The speculation on the matter of accuracy of how both methods deal with the motion data capturing takes into consideration most external factors as well as internal features of the techniques.

Part III

The Research

3.1 Human body and its 3D graphical representation

It is impossible to start working with humanoid shape without research on human body. Deep medical analysis is not needed, but it is essential to outline some most important points.

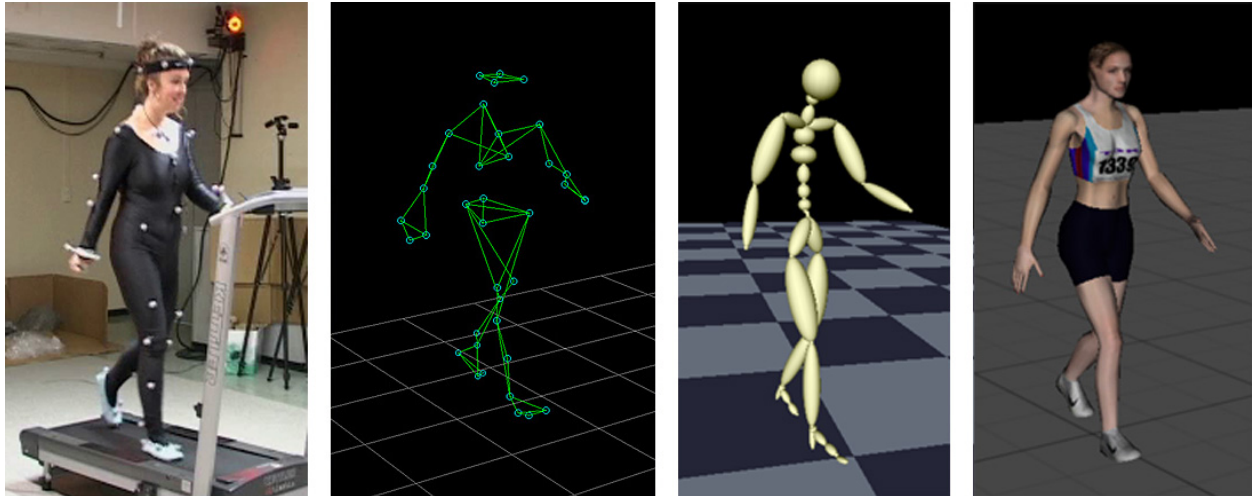
Firstly, the skeleton is a fundamental structure which is covered with muscles, skin, fat, etc. The appearance of the body form is influenced by it. Skeleton strongly affects the outer body contours. [2]

Secondly, the motion of the real human body comes from muscles contracting, stretching and moving the bones. 3D model of a human being to the contrary requires bones as a moving power. The bones cause all model shape's deformations. Such a structure is easier to setup and maintain.

When it comes to joints in human body, their main role is to allow movement. Some joints have wider angle than another, yet all are coordinated by the nervous system. 3D character for standard animation has less joints than the real body. Maya provides with two kinds of joints: hinge joints and ball joints. Hinge joints rotate in one direction and are often placed in knees and elbows, ball joints have three directions freedom, which is why they work for the backbone and shoulders. [1] (see Attachments for skeleton depiction)

3.2 Motion capture

Motion capture is a method used in 3D digital animation to record and adjust real-life movement sequences to virtual characters. Firstly it was designed to solve serious issues connected to a necessity of realistic human motion simulation in such fields as medicine, law, education and sport. Since it was by far more precise than keyframe animation, motion capture became involved in entertainment. Although, the cost of the hardware, additional equipment and software is measured as high, the amount of products based on the technology would probably increase. It has also become an essential part of Augmented reality concept, when the user's gestures are tracked and recognized in order to activate certain function of the 'hidden' application.



Picture 3.2: Motion tracking

Motion capture gives more accurate results when it comes to animating human-like characters. The method works only for the motion bound by laws of physics. For example, center of gravity movement is hard to realize with standard 3D animation techniques. On the other hand, motion capture is limited to realism. The model is required to have the same proportions as the actor to avoid the disturbing artefacts like reallocation of model's body parts. The action sequence involving other models might also need additional editing.

3.2.1 Technology

There are many approaches for tracking motion precisely and accurately. In general, all of these techniques fall into two major categories: optical and non-optical systems.

3.2.1.1 Optical MoCap

Optical motion capture technology is based on principle of measuring 3D tracks of sources attached to the actor's body. [7] Cameras serve as sensors, markers represent sources. Acquired data in most cases needs post processing to increase its usability, which means higher cost due to the additional operator's service. Basic optical system setup involves a single computer that controls digital CCD (Charge-Coupled Device) cameras that work on emphasis of photoelectric effect. Usually 60 samples per second rate is sufficient for capturing average human motions. Optical systems use from 8 to 32 cameras, each one can capture numerous images. All cameras require beforehand calibration in order to capture the desired object with specified in advance parameters. Although, larger amount of cameras has better precision, it also increases post processing time.

Most optical systems were produced for medicine, but found application in other fields such as

sport, filmmaking or video game industry.

Capturing process

Right after cameras' setup and calibration the software would respond to markers, ignoring the rest of the scene. Spherical form of the markers enables the camera to identify the light spot of each visible source. Then the software analyzes 2D images received from all cameras and, according to this information, counts the 3D coordinates of each point. Generally, several computer vision methods are combined to extract, recognize and capture the markers. After the image processing of 2D coordinates usually follows the stereo matching of the marker sets. Final three-dimensional Cartesian values represent a combination of personal and relative to cameras marker's coordinates. To distinguish one marker from another when filming the sequence, the software assigns an ID for each of them. The IDs are memorized until they are gone from sight of the cameras. If any marker is covered for a period of time, its ID happens to be reassigned. However, some modern software has the function of 'recalling' lost markers, which reduces the time operator spends on cleaning and editing parts of the tracks. Usually the epipolar geometry and stereo correspondence are used for the matching of markers. However, the mistakes may occur during this process, so the additional metrics, patterns, heuristics and algorithms implementing those are required. Otherwise the identification of markers would generate expensive combinatory operations.

Finally, the file containing the sequence of markers' global coordinates over time is saved and ready for the cleaning and fixing up process. However, the data does not form hierarchy, so further software assistance is required to gain a proper skeleton that would be used for animation. [8]

3.2.1.2 Non-optical MoCap

Non-optical motion capture systems use flex or inertial sensors, which focus on measuring flexion, acceleration and rotation. The emphasis is put on how human body moves, for example, how actor's joints actually bend, in order to summarise the motion data. Non-optical MoCap does not require complicated computer vision technology. On the other hand, it is constrained by being related to the 'last position', which is why scenes with more than two actors are difficult to shoot. [4]

Possible sensors would be potentiometers, powered goniometers, accelerometers and gyroscopes, all working with real human joints as sources.

3.2.2 Types of motion tracking systems

3.2.2.1 Passive optical system

So-called passive markers with retroreflective coating are attached to an actor's skin or special suit. Markers reflect the infrared light (IR) transmitted from near the camera lenses high frequency stroboscopes. Reflected IR light is used to define marker's position and send it back to the camera. It creates less visual disturbance for the user. [3]

The main problem of the system is how time-consuming it may be. Usually large amount of time is required to place the markers on the actor. What is more, fast enough movements or when the markers are too close to one another causes data confusion, which sometimes needs to be fixed by operator.

3.2.2.2 Active optical system

Instead of reflective markers the system manipulates with light-emitting diodes that have integrated processors and radio synchronization. Each LED gets an identifier (ID), which allows to distinguish the marker and even recall one when it reappears after being hidden from cameras. The cameras' shutter speed is synchronized with LEDs to provide capturing. [8] The rest is similar to passive system.

Active optical system needs an additional LEDs controller which is attached directly to the actor and may interfere with his movements. Active markers are less vulnerable to disturbances, such as external light, but also require own power supply.

3.2.2.3 Optical markerless motion capture

The very special type of MoCap systems which is seen as the future of the industry still requires a handful of research. For example, developers from IPISoft work on the problem as well as Organic Motion and Xsens. The system consists of single or multiple cameras and software that uses image processing algorithms and computer vision to read skeleton animation from recorded video.

Basically, the contour of the human-like shape is extracted from the images received from cameras. The contour is calculated as a difference between the current frame with an actor and the initial frame of the empty scene. Actor's pose and shape are observed, divided into parts and tracked down at every time. [10]

3.2.2.4 Magnetic system

Magnetic system came from military aircraft. The principle was used for pilots' helmet-mounted displays. Magnets work as markers and receivers - as cameras. Sensors placed on the body move in specially generated low-frequency electromagnetic field. Software receives the file containing both positions in Cartesian coordinates and orientation. [8]

Magnetic system can be vulnerable to magnetic and electrical interference from surrounding metal objects like wires. Sensors' response varies from each one's position in capture area. Operational space is also more limited in comparison to any optical system. [3] At the same time it is possible to capture more than two actors performing simultaneously and even interacting with other objects.

3.2.2.5 Mechanical system

Mechanical motion capture system follows the joints' bending directly. For this issue the actor puts on mechanical mocap structure which repeats his every movement. The gathered data describes all joints' angles. What is more, the system allows real-time data compilation. [8]

On the other hand, the actor's moving possibilities are restricted by a quite fragile mechanical skeleton. For example, one can not perform sequences involving falling. Mechanical system does not support tight co-working of two or more actors like dancing or fighting as well as the facial expressions capturing. The sampling rate is also lower and sometimes the additional system to compute global translations is required.

3.2.2.6 Inertial system

Markers or magnets in inertial system are replaced by miniature gyroscopes and inertial sensors. This way the data contains not only the position of the sensor, but also its pitch angle which is translated to software skeleton model.

The main advantage of the system is that it covers large capture area. However, the additional mini-system (magnetic or optical) is required to specify the global position of the actor in order to avoid one's image from 'floating'.

3.2.3 Data formats

VRlab provided with an optical system and passive markers. The software for motion tracking management was Motive from Optitrack which exports .csv and .c3d files, readable by Motionbuilder.

Additional Maya-plugin was installed to support .fbx file with animated character. Another popular data format widely used in motion tracking is .bvh.

3.2.3.1 C3D

C3D format was created for biomechanics research facilities, but was then developed by Vicon for film and computer games animation. File with C3D extension is binary and it stores all the data in one place. The file has three parts: Header, Parameter and Data, in average 512-byte for each section. C3D file may contain additional information like actor's characteristics. [8]

3.2.3.2 Comma-Separated Values or CSV

CSV is format that stores data table as plain text. Each row is equal to a table row. Columns are divided using comma or any other suitable delimiter. CSV standard is not strict and also quite portable. More information can be found in [12].

3.2.3.3 BVH

Biovision Hierarchical data format was developed by Biovision to describe the data as joints forming a skeleton for motion capture animation. BVH file contains hierarchy and data stream sections. First one is represented in blocks starting from the joint specified as 'root' and shapes a tree graph with the translation data included. Hierarchy allows to trace child's position in relation to the parent node, however, it is impossible to learn what was the orientation of the node when it was initialized. Second part, Motion, contains rotation data. Example of the .bvh structure can be found in [8].

3.2.3.4 FBX

FBX (as in FilmBox) is a 3D file format developed by Kaydara for MotionBuilder. Since 2006 it belongs to Autodesk, which means that the .fbx extension is supported by Maya and other company's products for 3D design. FBX has two versions: ASCII text version and binary one. Text version represents each node's ID, parameters and list of related nodes. Autodesk provides the users with free open source SDK library. [13]

3.3 Rigging

Rigging is a process of preparing the 3D character for the animation stage. It involves building of a virtual skeleton and placing it inside the model. The skeleton consists of simplified bones and joints and presents a hierarchy of those elements, meaning that each bone translation or rotation would influence all dependent nodes as well. Skeletal animation is convenient, because it provides with the ability to manipulate a vast number of figure's parts like eyes, muscles, hands, etc., using only a few controllers, bones and their characteristics. [1]

Part IV

The Analysis

4.1 The software

There is a great variety of software products designed for 3D modeling and animation. Students of Web and Multimedia in Czech Technical University learn how to work with quite a few. Although Blender is a freely available and decent program, when it comes to working with the details personal experience gained from A7B36KMA course suggests Maya of Autodesk as a better option. Maya has become a 3D graphical standard in filmmaking and television. Aside from being flexible and powerful tool, Maya is quite favourite among side developers due to its openness. It can be converted into a version that would satisfy the requirements of big studios.

It was decided to use the release of 2012, because it is stable and reliable. Previous experience with the newer versions of Maya was not as positive.

Besides Maya 2012 64-bit which was used for character modeling and final rendering of the animation sequences, the following programs have been exploited for the thesis:

- Autodesk MotionBuilder 2014 for application of MoCap data
- MotiveTracker for camera calibration maintenance, data management, cleaning and editing
- Kostilam, additional applications and libraries provided by supervisor Ing. David Sedláček for markerless tracking
- FFmpeg - complete, cross-platform solution to record, convert and stream audio and video [18] was used for video processing during markerless motion tracking

4.2 Model

Starting model was specified as humanoid, which means human-like form of any precision. Still it is essential to reduce possible artifacts appearance range. The less similarity exists between the model and the actor, the more errors may occur. Which is why the mesh was created using polygonal modeling. The character is very basic, does not strictly belong to any gender, has no fingers and no facial features. The model would be quite easy to change according to the participant's parameters. It also does not have textures, only basic materials so that the attention would be focused on the motion.

The used rigging technique is also very common for human-like models. It is actually the default technique and is supported by products designed for motion capture. Since both Maya and MotionBuilder use the same standard for skeletal definition, it would be easy to identify the parts of it in both programs. However, some occlusions are expected during the exporting of .fbx files, because the used versions of Autodesk products do not match.

It is expected that additional editing would be required to lessen the model's skin deformations as the real human motions might be quite hasty.

4.3 Testing the technology

The best way to test the main functions of the provided hardware and software is to go through all the phases of actual motion capturing process and fulfill the tasks at hand. It will also allow to describe how the data is collected, edited and applied on the model more accurately, to experience the problems that MoCap operators face and try to solve them.

Each session in VRlab would involve shooting five independent sequences. Presumably, four sessions would suffice the cause of this thesis. The very first would be dedicated to studying how to work with the cameras and Motive:Tracker without the actor. The idea is to observe the way the cameras deal with simple translations of a few markers and learn the basics of editing in the Optitrack software. The second session would be with an actor in a special MoCap suit and with markers attached to the body. For the third session, the cameras would be moved according to observations made during testing with the initial installation and recalibrated. Then the motion sequences would be repeated and recorded. The last session would be markerless motion tracking.

The motion sequences would be 'jumping jacks', running, fighting or boxing-like moves, dancing and improvisation of an actor. If required, additional records would be made. The sequences and their performance are described further in the text.

4.4 Preparations with the actor

Model's rig depends on markers due to the fact that sensors provide the important data for every joint. In passive optical systems correct and sufficient markers' placing is crucial for collecting the accurate data. On the other hand, larger amount of markers on the actor would mean extended cleaning time during the editing stage. Markers should also not be placed too close to one another or they might be swapped.

To avoid unexpected translation the world axis would be chosen in advance and would be identical for both the character in Maya and the image of the actor's figure inside the capture area.

One of the cameras would be pointed out as main. The actor would start and finish the sequence faced to this camera. This way it would be easier to control the sequence flowing.

In order to gain better knowledge on the technology, it is essential to run a group of different tests by trying out slower and faster motion sequences as various as the capture area would allow. The actor would also be instructed beforehand on the matter of action sequences that would be performed as these should stay the same from session to session.

4.5 Applying the data on the rig

Standard Maya package does not contain .fbx import and export, which is why free Autodesk FBX Plug-in would be required. The plug-in is easy to install, so the process of further file transferring should not cause any troubles.

In order to resolve some problematic issues with the character's rig and motion capture rig non matching, Maya of 2011 and its later versions provides with HumanIK middleware, which contains retargeting function that serves to adapt animation for the concrete model. [13]

4.5.1 Motion Capture skeleton and Maya

The main idea behind successful motion capture data application on the character's rig is in making matching skeletons. This could be achieved by working with standard human skeleton structure featured in different literature like [1] or [8] and by accuracy of marker's positioning. If two skeletons match, the translation and rotation data from one needs to be attached to every equivalent joint in the other.

Although, the manuals advice to avoid such collisions and disturbances as dislocation of the model's body parts after the MoCap data application by having an actor with the similar complexion as the 3D model, it might not be always possible. The only tactics here is to alter the character if necessary rather than change the participant.

4.5 Markerless shooting

Due to the fact that markerless motion capturing is a developing concept, the application stage would be eliminated for now. On the other hand, thanks to software provided by supervisor, the records would be processed to generate the batch files containing motion data sequences available for study, evaluation and any further related research work. The comparison of two motion tracking techniques explored in the thesis would be mainly based on quality of produced results.

4.6 Cameras' positioning

The hypothesis for the matter of cameras' influence on the process is following: the positioning has a great impact on the condition of markers' tracks. To prove or deny the statement the cameras are going to be translated somehow. However, whether or not the positioning influences the markerless MoCap is uncertain, because the external information on the subject is limited. Since the studied optical motion tracking without sensors uses video based approach, it was assumed that the way cameras were installed would affect the produced actor's image.

Part V

The Workflow

5.1 Stage 0: Modeling and Rigging

The 3D model of Smiley was created while working on the project. The polygon cube had additional edges inserted and was stretched and extruded to get the result that can be seen on pictures in Attachments section. The model represents a basic human-like shape, reminding a skeleton. Such a design allows wider range of further alterations in case of adapting the model to the proportions of the actor. In the attached Maya files it comes as rough polygonal mesh, on which back and forth smooth operation had been applied to track down, how the skin deforms when the character's pose changes.

Although, the 3D model was produced with male type of figure in mind, it had to be adjusted to a slim female participant.

5.1.1 Creating the rig

Using the Joint Tool, the joints were put into the legs of the character, starting from the root, then the hip and down to toe. Each joint was renamed and labeled properly. Automatic Control rig provided by Maya HumanIK was used as it suffices MotionBuilder the most. Model with Control rig comes separately though, since the MoCap moves are already bound to physics of human body motion, it is not needed to specify further constraints for such a basic type of model.

As a result the skeleton tree structure was made with joints as nodes. Body parts are identifiable and any leaf's translation would not reallocate its parent-node.

Previous version was also included and can be found on the disc.

5.1.2 Placing controls (prev. model version)

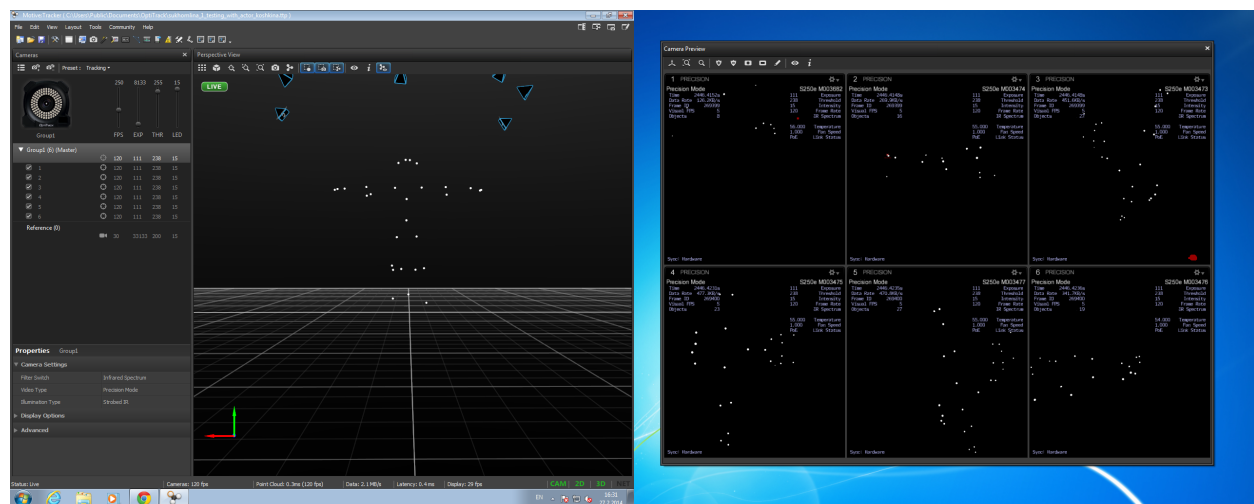
Controls serve as helpers to the animator in keyframe 3D animation. Those are curves forming practically any convenient shape and having their own hierarchy. By Assisting controls for knees, elbows, backbone, palms and feet were created for smoother operation. These should be sufficient for testing and further fixing and editing when the motion capture file would be applied.

5.2 Stages 1 and 2: Recording and Editing

Motion capturing sessions took place in VRlab of Electrical Engineering Faculty, CGI department. First session, as mentioned before, was for the sake of getting to know the equipment, software and hardware and was conducted by supervisor. Then the author of the thesis experimented with the markers for a while. The built-in timer of MotiveTracker was used to delay the start of the recording. That is how the author could take the position inside the capture area with the three-marker bracelet on the palm of the hand and observe how the position is changed on six small camera view projections of the Calibrate layout. The cameras were in the initial position at the ceiling. It was also noticed that rigid body 'create' and 'control' functions of the program allow to distinguish the relevant coordinates of each marker belonging to the specified group from another one in the same set. It was decided to try rigid bodies later in the editing phase.

5.2.1 Session 1

The team consisted of the author, who directed the participant and managed the recording at the VRlab computer, and the KME FEL student Ekaterina Koshkina as the actor. Cameras had been calibrated in advance by the supervisor. After the actor put on the black MoCap suit borrowed from Ing. Zdeněk Trávníček (IIM FEL), the markers were attached to the shoulders, arms, hands, chest, hips and legs, including 4-markers headband. When the external light was removed, Optitrack cameras received an image of markers. With a higher threshold and masking the disturbing effects were insignificant. The recording was committed with cameras set to 120 fps by default.



Picture 5.2.1: Session 1 - Motive layout

The actor was asked to perform predefined sequences of movements, each time starting and finishing with the basic T-pose which is visible on the picture 5.2.1 above.

5.2.1.1 Editing

Five sequences were recorded including one extra as a backup. The next step was to clean and edit the collected data which was also the longest and the most time consuming phase, as it required operator to manually organize the marker related data. By pressing right mouse button and choosing Trajectorize option, the calculation of points' positions in three-dimensional space for each frame was enabled. The separate trajectories were combined together.

When the operation finished, a list of unlabeled markers appeared. Usually, the markers predefined as rigid bodies are maintained automatically. However, during the shoot the markers of the headband specified as rigid body (meaning that their position is definite) swapped, causing unwanted deformation of the model's head during application. That is why it was considered not to use the rigid bodies in the next session.

For each take several Marker Sets were created. Those were named as Head, Chest, Arm Left and Right, Leg Left and Right and Hips - to distinguish each body part. By activating the Labeling view, starts the actual 'puzzle solving'. One of the main drawbacks of Optical system with passive markers is how they get lost during the motion recording. Especially if the actor performs sequences where he needs to cross arms or legs, for instance, running. Markers which got covered from sight of the cameras are then renamed and require to be found in the list among other unlabeled and parasite-markers. In case of fast motion performances, a whole lot of such pieces is received. The process of restoring the track is quite simple, though. Operator follows the record to track down when the actual loss has happened (marker disappears and then returns back coloured as unlabeled with the new ID). The failure is represented in the graph of marker's trajectory by a gap. All markers with such tears are pointed with G next to the Marker-ID sign. A trajectory is restored by finding the pieces and stitching them together. If the bigger part of information was lost, Motive:Tracker provides with tools like manual drawing of the curve, filling in the gaps based on interpolation, smoothing by certain frequencies, etc., all of which were used during the editing stage.

When the count of G's was reduced or eliminated, each take was exported as .csv and .c3d file.

5.2.1.2 The outcomes of Session 1

After the cleaning and fixing procedures, five files were exported for further application in MotionBuilder. The sequence of running around the scene was damaged beyond repair, because knee markers swapped and then were gone from camera view for good half of the recording. So the sequence had to be replaced with an extra take: running on one place. Two out of five exported files contained gaps which did not create any major problems during Application stage, though. Actually, the chosen editing strategy with the help of rigid bodies was the cause of model's shape breaking and twitching on rendered videos. Actor's flouncing, which is natural for human body behaviour, was also interpreted as unsmooth 'jumpy' shifting in Application phase.

The author noted that actor's natural movements were not the only cause for partitioned trajectories and multiple tears. The losses were also at the ends and sometimes at the beginnings of the recordings when the actor was just standing in a T-pose. Either left or right half of the body has always been in worse state than the other. Markers on arms and legs used to disappear during twists and turns, when they got between two neighboring cameras on sides with no cover from behind.

The decision was made to cover the so-called 'blind spot' at the back of the room and distribute the rest of the cameras in a circle. More markers were also adjusted to define the spine, collarbone, hands and feet as a precaution.

All in all, inexperienced operator and actor could also be added to the list of issues that lead to the repetition of the experiment.

5.2.2 Session 2

5.2.2.1 Calibrating

With the great help from supervisor Ing. David Sedláček it became possible to rearrange cameras. The positions can be found in Attachments. ([Session 2, alternative camera setup:](#))

Before the actual MoCap recording the system had to be calibrated which is a normal practise after every conducted change and in case the cameras have not been used for a long time. Calibration requires elimination of the external and internal light sources except for IR light emitted by cameras. Another condition is to exclude the objects reflecting infrared spectrum. MotiveTracker enables masking of parasite reflections.

The calibration involves walking in the capture area and waving with a special tool called the Wand. The camera views depict the coloured traces left by markers attached to it. The process should not be paused until the views are filled, and the quality is measured as High. Calculation of the camera parameters, their positions and orientation is iterative and can take around 5 minutes, although the results are usually available right away.

The last but not least is to define the 3D coordinate system. A special triangle tool is placed upon the desirable spot inside the capture area. It defines the starting position of the scene. That calibration file could be used until the next reconfiguration, which took place in Session 3.

5.2.2.2 The outcomes of Session 2

Cameras' influence

Just like in previous session five data files have been exported. Judging from the editing and cleaning experience, slight positive changes were noted, especially for the legs and the upper body marker sets. It allowed to shoot running sequence, which was damaged in Session 1. However trajectories of sensors that had been attached to arms and hands were still hard to follow and restore. Due to the fact that the number of markers had increased, it was possible to gain necessary translation data required for right animation.

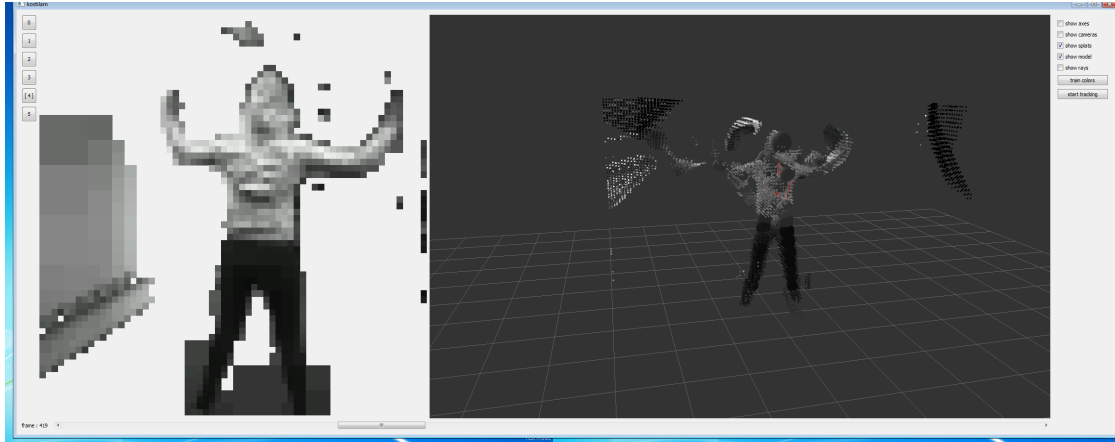
Markerless MoCap: why it failed first time

Session 2 was supposed to be divided into two parts: motion tracking with and without markers. Actor was asked to remove the markers and to perform the same motion sequences as before, however, due to insufficient lighting, the recorded videos were exported as darkened .avi files that would not have been recognized by the processing software. Therefore, Session 3 was arranged.

Altogether, Session 2 provided with some answers for the task questions of the thesis and was evaluated as successful.

5.2.3 Session 3

The last motion tracking experiment had the actor and the operator exchange roles due to an injury of the first. The author of this thesis performed simplified motion sequences that were recorded at 60fps in a normal camera mode in a well lighted VRlab. The videos were exported and then processed using Kostilam and predefined 6 camera calibrated positions and orientations. The videos from all cameras were masked. By defining the beginning and end of each motion sequence and the empty frame as background, the tracking operations were enabled in a generated batch file.



Picture 5.2.3: Kostilam, 'jumping jack' sequence

The program is able to recognize the participant's figure by precision and by knowing camera parameters. A 3D skeletal structure made of interconnected spheres in the center of the scene has to be placed inside the figure to start following the recorded movements. Due to various reasons, such as poor camera setup, non tight-fitting clothes of the actor, complex background with multiple objects around the capture area, etc., the motion sometimes seems blurred and unfocused. The skeleton cannot catch up with the actor and has to be moved manually. It also floats above the ground plane, moving up and down Y-axis. Perhaps it requires extra ground definition constraint.

Considering faster motion sequences, like running, the results produced by motion capture with passive markers became clear after editing. Markerless processing, on the other hand, created particle clouds, so it seems troublesome to track legs' translations.

Lighter background produced additional particles above the figure, especially in the last sequence 'improvisation', which operated with floor touching body poses. Presumable, darker upper body clothes could solve this problem.

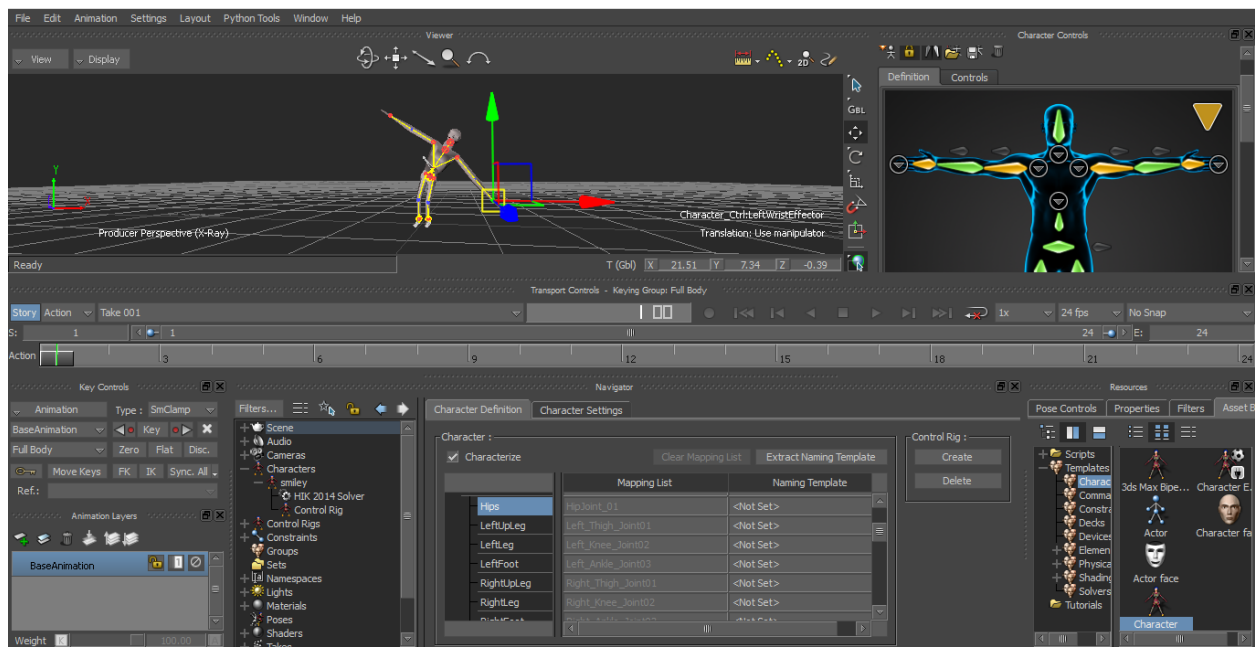
Concluding on the matter, markerless technique might become a welcoming alternative, since it would not require so much manual editing. Even now, just from observing the way the skeleton is influenced by particles of the silhouette in real time tracking, it is clear that markerless MoCap would someday produce better results than the already classical passive markers system. Although, it is hard to tell whether the actual application on a 3D model is going to be smooth, the capturing and editing stages gave a good impression.

5.3 Stage 3: Application

MotionBuilder is a convenient program for motion data application. It operates with .fbx files and recognizes exported Motive:Tracker motion capture data packages. The interface is quite similar to Maya and is quite flexible despite the amount of windows opened at once.

First step was to import .csv or .c3d files with markers' positions. Then the Actor was added to a scene, which is a special MotionBuilder puppet-like structure with empty slots on joints positions that can be filled with markers data by simple drag and drop from scene navigator window to the editor. As soon as markers are placed inside the joints, and the actor is bend a little to fit in their cloud at the scene, active mode can be chosen. The Actor would repeat the prescribed motion.

The next step is to import model from Maya that should be animated. This has to be done carefully, since sometimes the Actor tends to break, so the active regime is disabled for a while. For the first time trying to define the Smiley character in MotionBuilder by assigning bones from the model's skeleton to the ones in Character definition window on the right, the root joint was missed which caused dislocation of the upper and lower body of the model after opening the scene back in Maya for rendering. The problem was solved by skeletal rearrangement in model version 2.



Picture 5.3: MotionBuilder

When all basic parts of the skeleton are defined, the source for the character was chosen as 'Actor' and then biped for humanoid type of models. By confirming the choice for the animation source, the character starts to perform the same motion sequence as the Actor before.

That is not all, however, the animation still needs to be attached to the character directly so it can be transferred to Maya. From the Character settings menu the Plot Character button was clicked. MotionBuilder offers two ways of animation binding. One is to adjust it to the control rig and the other one - to the joints of the skeleton. The second option was preferable. Finally, the .fbx file with both model and MoCap data is exported and ready for further manipulations in Maya.

5.4 Stage 4: Rendering

After the scene with animated 3D model has been opened in Maya it was still early for rendering. Interactive Skin Binding used while creating Smiley version 1 required additional editing like weight-painting, since the moves cause skin deformation. The mismatches are also inevitable, since the participant and the model are not identical, not to mention the failures in markers' data.

Version 2 manipulated with Smooth Skin Binding option, because Interactive Skin Binding was too complex to handle for such a basic humanoid model. After little editing of Skin weights, the scenes were finally processed with Maya Batch Render.

The results can be found on DVD as Maya files and .avi videos.

Part VI

The Conclusion

The requested tasks have been fulfilled. Recently purchased VRlab hardware and software have been studied and operated to recreate the motion capture data processing pipeline from recording to application on the model designed specifically for this purpose. Source files from different phases describe the process as a whole and are available for further implementation.

On the provided example of five takes produced using both methods: with and without markers, the advantages and disadvantages are quite clear. The point of the author on the matter of markerless MoCap is that it should be developed for the benefit of all involved in the field.

The thesis is written on the emphasis of Radek Smetana's Humanoid Model work, but looks at the subject from a different angle. Generally, it was more related to experimenting with the wide range of functions and testing older against newer approaches in MoCap. The main goal, though, was to assist in adapting this concrete technology to VRlab environment.

Motion tracking data records can be used for other projects as well as for study purposes. Observations concerning the camera positioning, markers adjustment, sessions management and MotionBuilder might also be helpful for novices.

The results received using Optical MoCap system with passive markers are generally fine with some occasional gaps and failures. Although, some of these tears in trajectories have been edited and the curves heavily filtered, it is still not recommended to rely on editing tools too much. Basically, because the idea behind this and other motion tracking techniques is to capture realism, not provide smooth data. The takes are usable for 3D animation, so the goal is reached.

Markerless MoCap data files have been generated as requested. Although, a bit broken, they contain the distinguishable motion sequences that can be followed and maybe even applied on a 3D character in future.

Bibliography

[1] MARAFFI, Chris. Maya Character Creation: Modeling and Animation Controls. New Readers Publishing, Berkeley, California, USA, 2004.

[2] RATNER, Peter. 3-D Human Modeling and Animation. John Wiley & Sons, Inc., Hoboken, New Jersey, USA, 2003.

[3] "RENDER.RU -> Статьи -> Эксклюзив -> Все о МОСАР." RENDER.RU -> Информационный ресурс по компьютерной графике и анимации | 3D, 2D галереи, уроки, все для CG художника. N.p., n.d. Web. 10 Jan. 2014.
http://www.render.ru/books/show_book.php?book_id=531

[4] Markerless Motion Capture System. N.p., n.d. Web. 10 Jan. 2014.
<http://www.organicmotion.com/applications/animation/>

[5] SMETANA, Radek. Model humanoida. Praha, 2011. Dostupné z:
<https://dip.felk.cvut.cz/browse/details.php?f=F3&d=K13139&y=2011&a=smetarad&t=bach>.
 Bakalářská práce. ČVUT. Vedoucí práce Ing. David Sedláček.

[6] OptiTrack manuals and documentation <http://www.naturalpoint.com/optitrack/products/s250e/>

[7] SILAGHI, M., R. PLANKERS, R. BOULIC, P. FUA and D. THALMANN. Local and global skeleton fitting techniques for optical motion capture. Proceedings CAPTECH'98. 1998, s. 26-40. <http://infoscience.epfl.ch/record/98855>

[8] MENACHE, Alberto. Understanding motion capture for computer animation and video games. 2. ed. San Diego, CA: Morgan Kaufmann, c2000, xiv, 238 p. ISBN 01-249-0630-3.

[9] BALLAN, Luca and Guido CORTELAZZO. Marker-less motion capture of skinned models in a four camera set-up using optical flow and silhouettes. 3DPVT 2008, Atlanta, GA, USA

[10] «Захват движения» от iPiSoft / Блог компании Intel / Хабрахабр. (n.d.). Retrieved from <http://habrahabr.ru/company/intel/blog/103718/>

[11] MEREDITH, M. and S. MADDOCK. Motion Capture File Formats Explained.

[12] CSV Comma Separated Value File Format - How To - Creativyst Explored, Designed, Delivered.(sm). (n.d.). Retrieved from <http://www.creativyst.com/Doc/Articles/CSV/CSV01.htm>

[13] <http://www.autodesk.com/products/>

[14] Риггинг 3D модели - Мир 3D. (n.d.). Retrieved from <http://www.mir3d.ru/learning/930/>

[15] <http://www.motioncapturedata.com/2009/08/mocap-file-importer-maya-plugin.html>

[16] <http://www.creativecrash.com/maya/script/bvh-importer-for-maya>

[17] HARTLEY, Richard a Andrew ZISSERMAN. *Multiple view geometry in computer vision*. 2nd ed. Cambridge: Cambridge University Press, 2003, 655 s. ISBN 05-215-4051-8.

[18] <http://www.ffmpeg.org/>

ATTACHMENTS

Abbreviations

3D - three-dimensional

actor - person who performs the motion capture sequence

fps - frames per second

MoCap - Motion Capture

A7B39KMA - Multimedia Applications Course

A7B39PRO - Semester Project

T-pose - initial position of an actor, reminding latin letter 'T'

VRlab - Virtual Reality laboratory of DCGI department, FEL, CTU

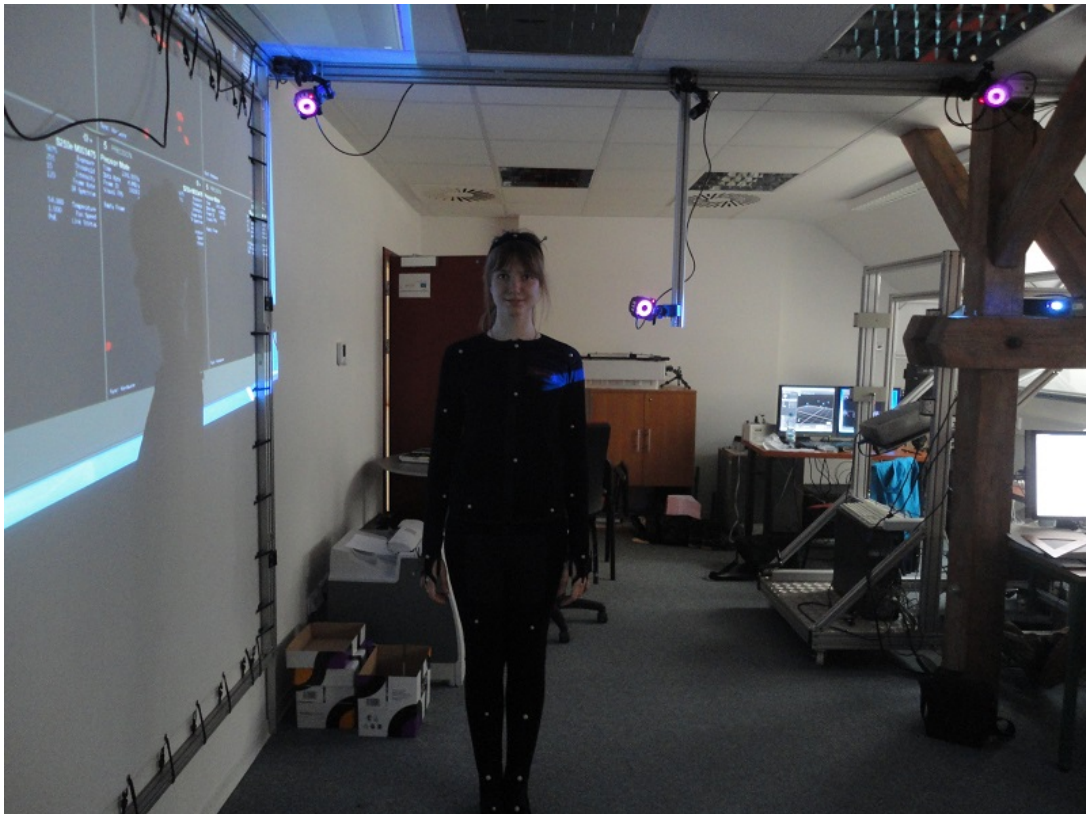
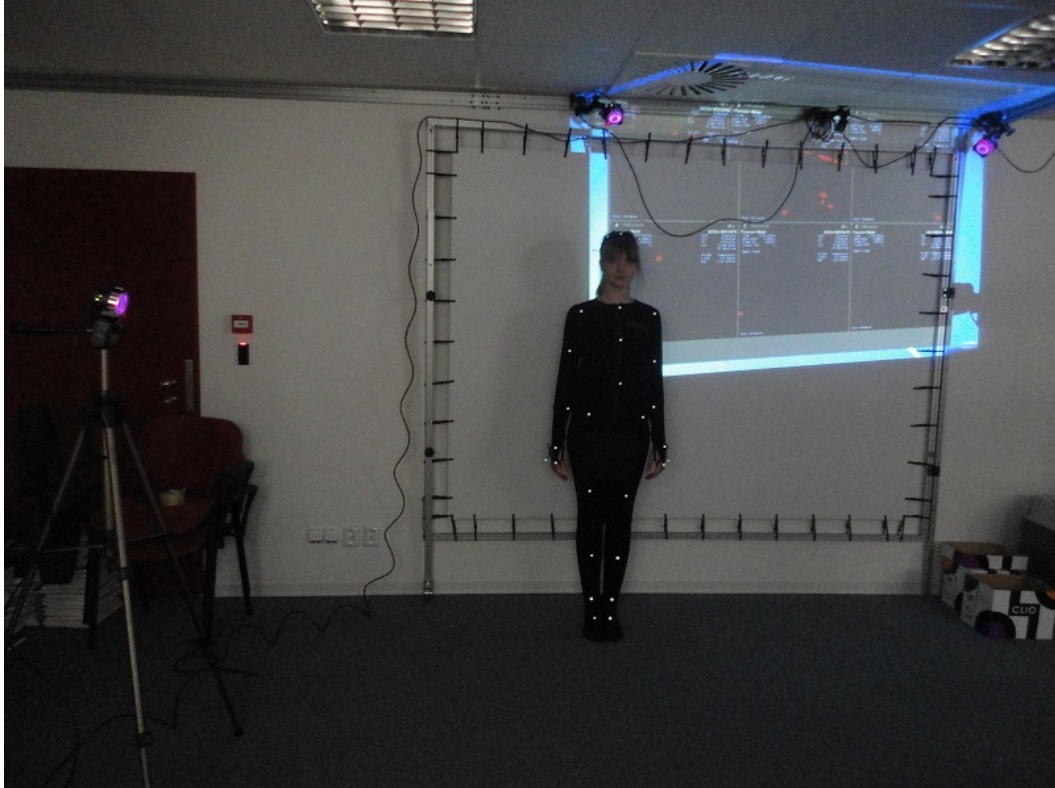
Pictures

Initial Camera setup:



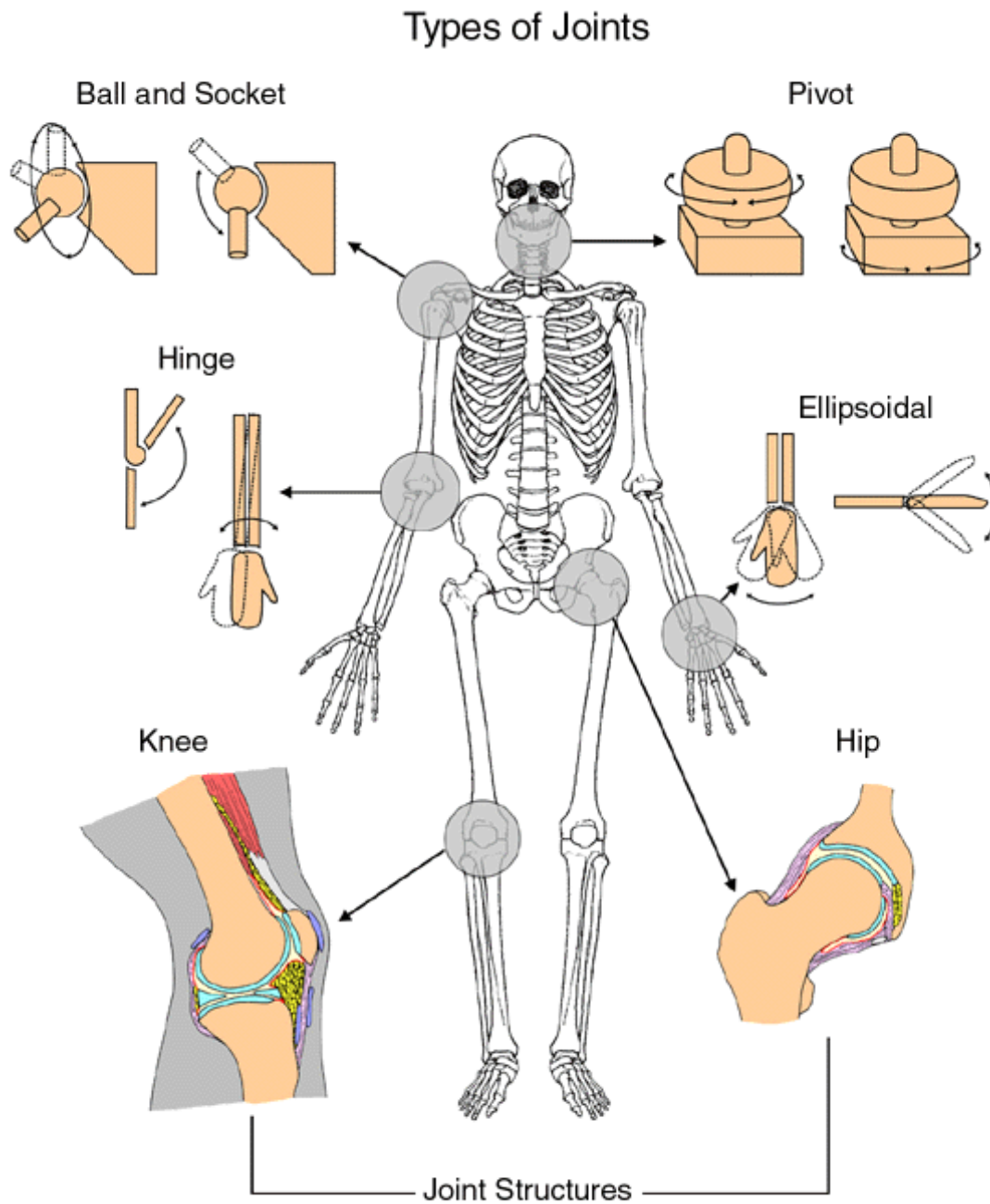
Session 2, alternative camera setup:





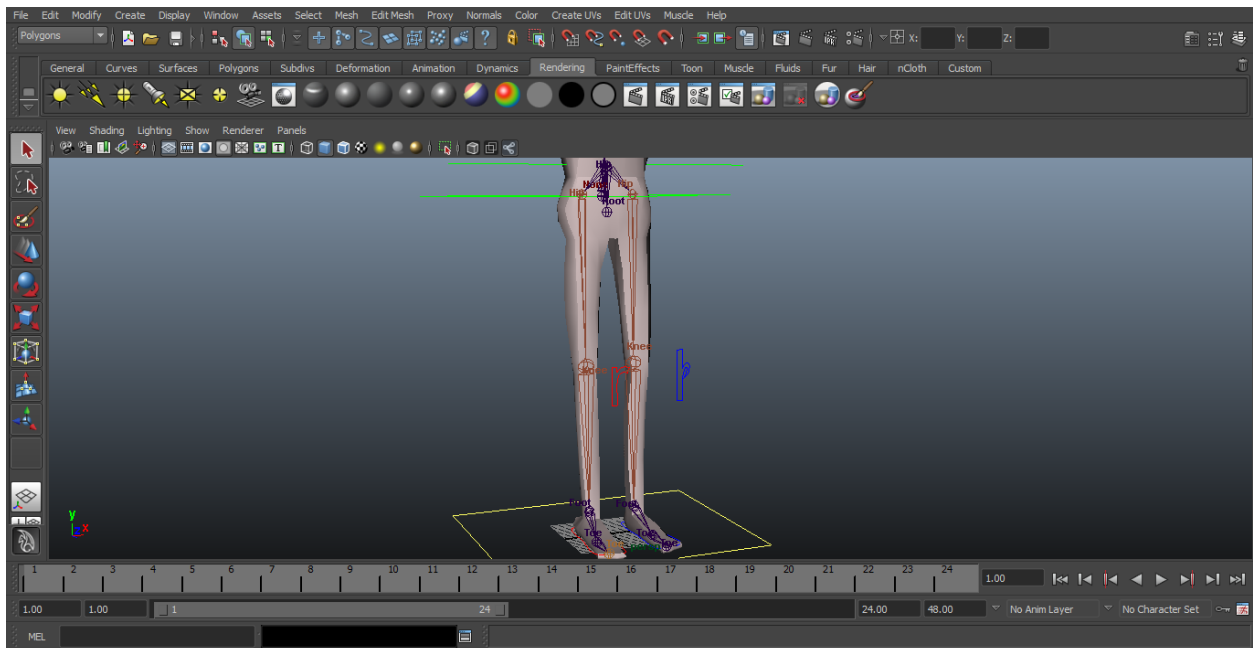
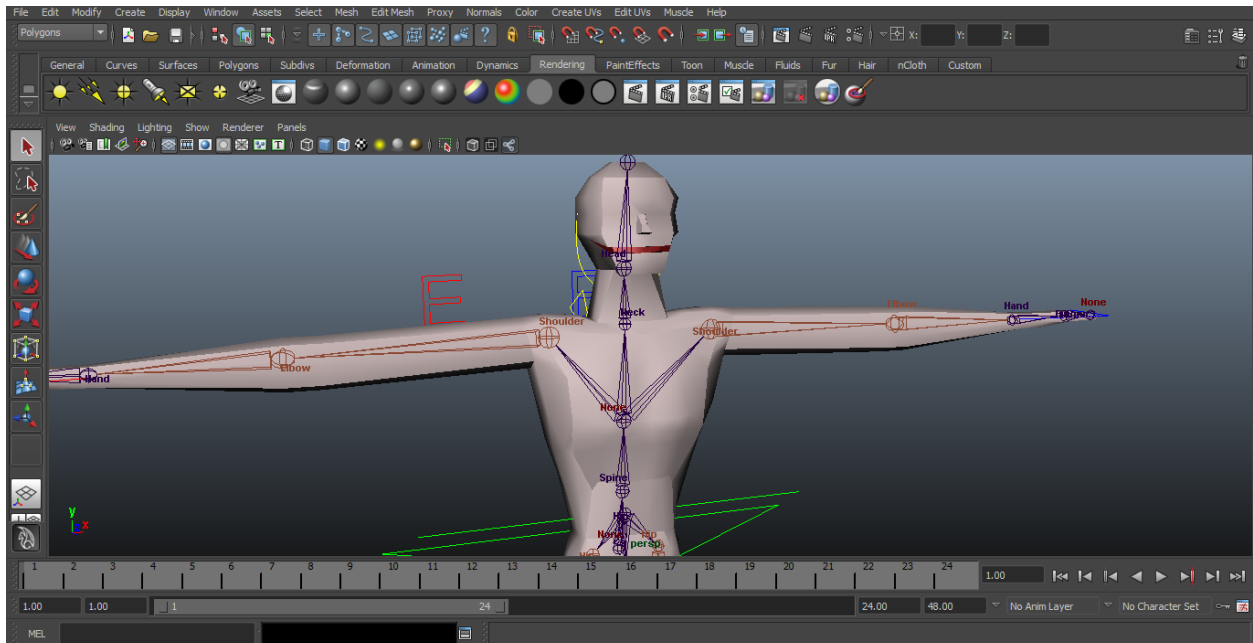


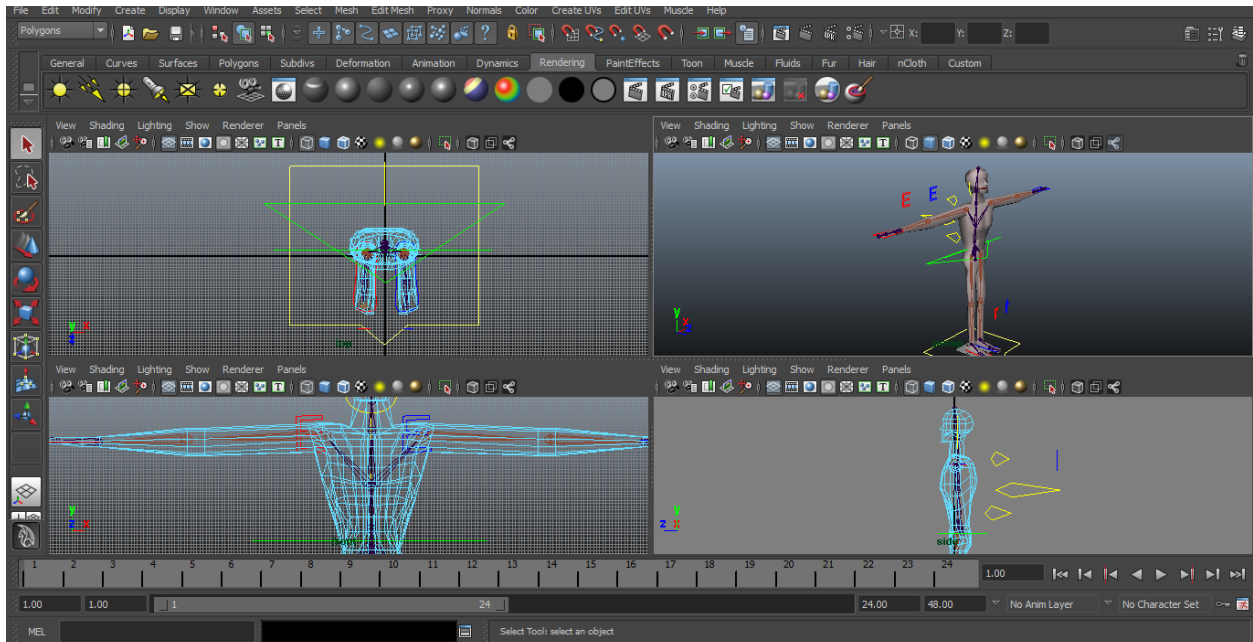
Joints in human skeleton:



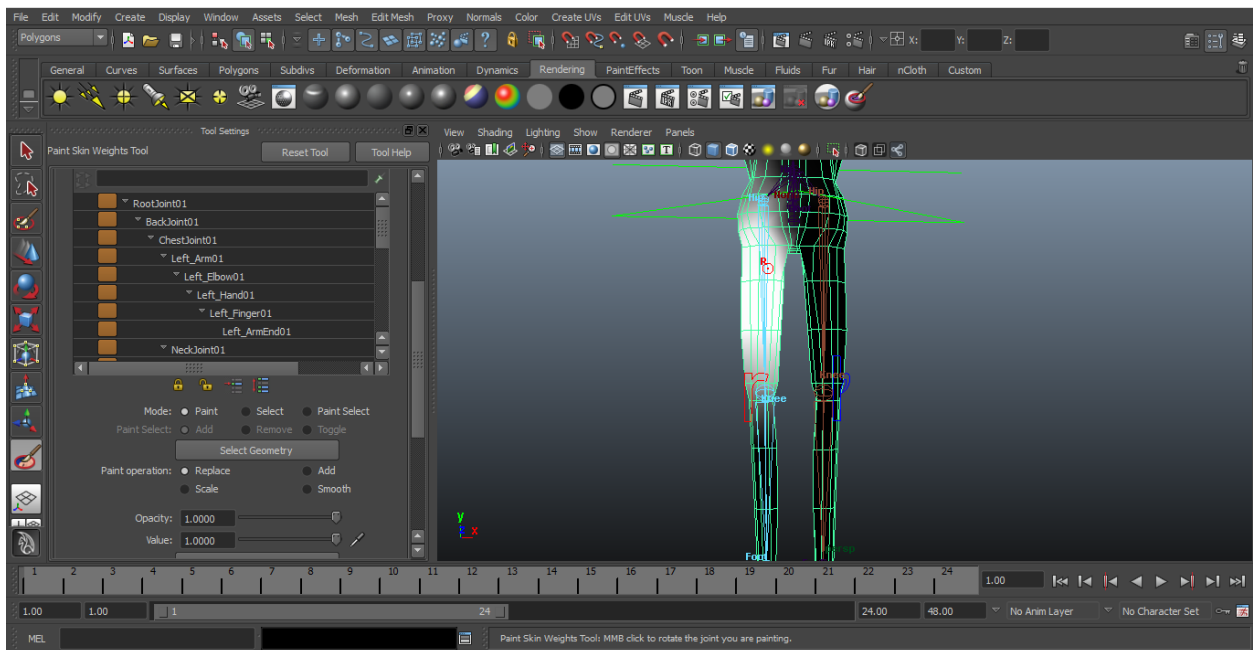
<http://www.aokainc.com/wp-content/uploads/2013/09/ball-hinge-joint1.jpg>

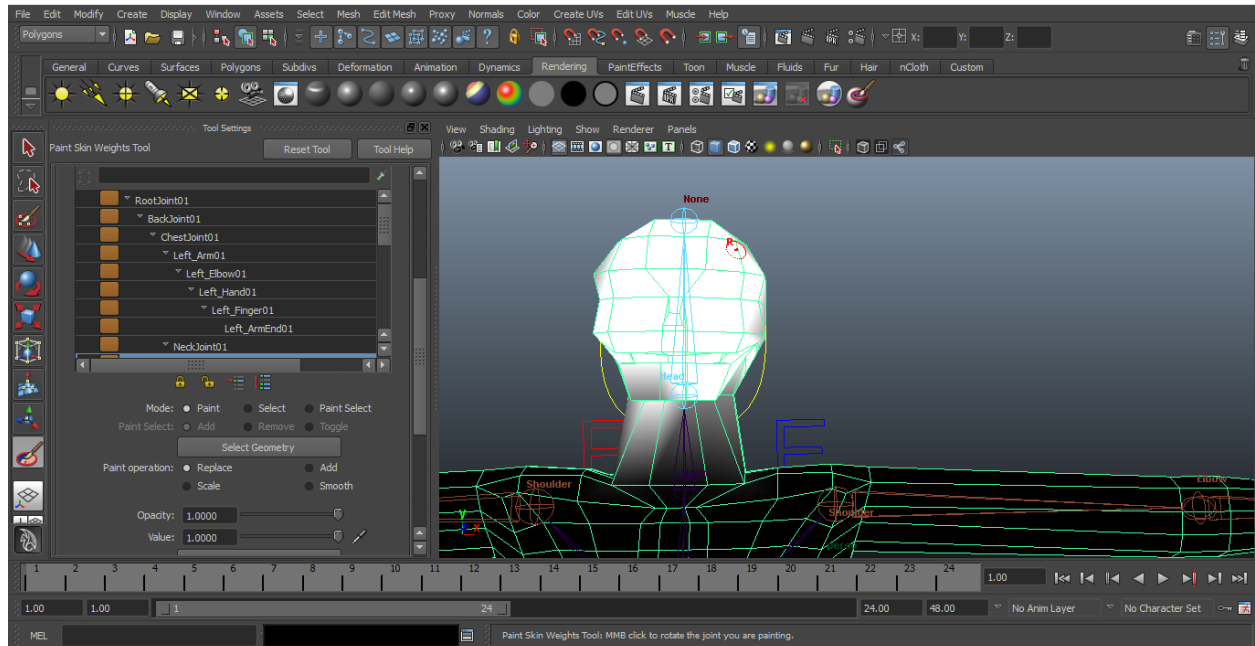
Rigging 'Smiley' version 1:





Skin weights painting example:





Camera calibration using the wand:



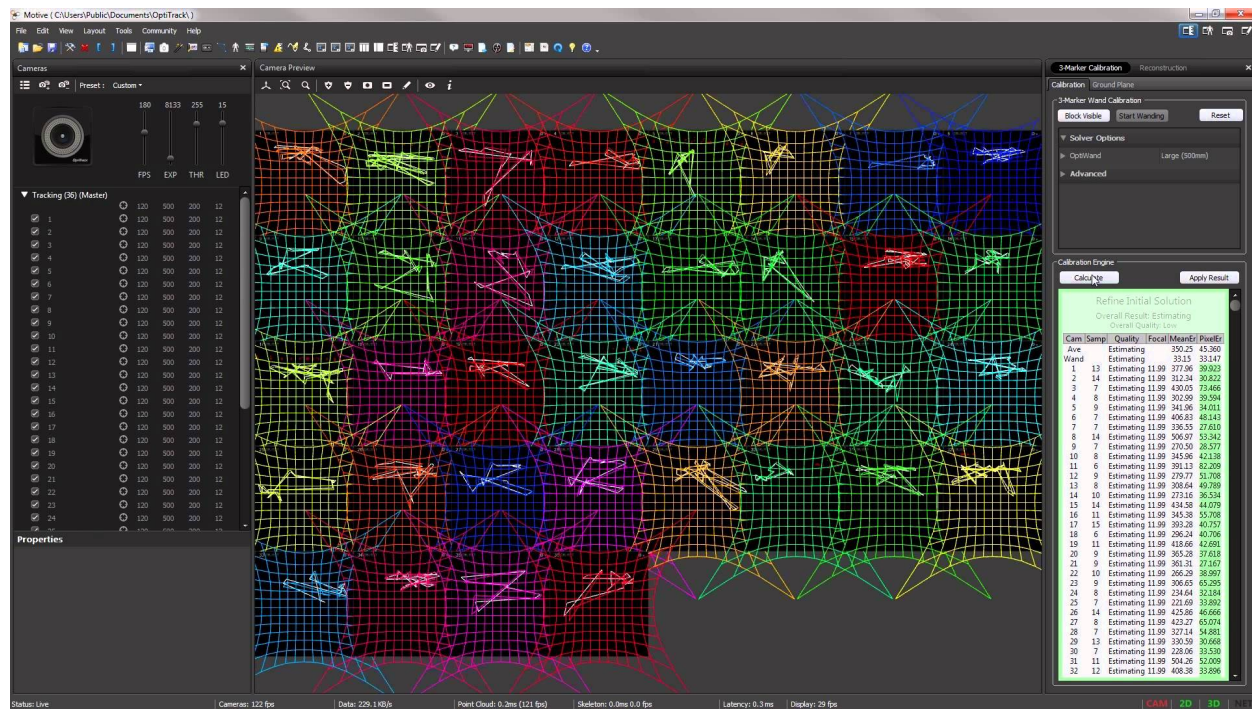
https://encrypted-tbn1.gstatic.com/images?q=tbn:AND9GcS6vcL0O05n6Db8GL8-itTkct_rvOHZ8asHyWAgIt6wk44N8wYZew

OptiTrack Calibration Square:



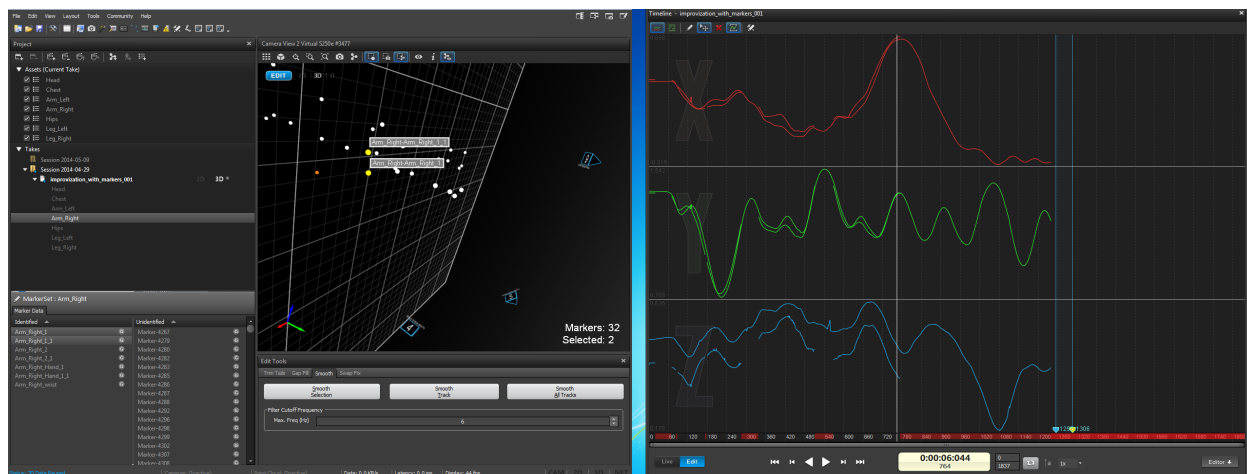
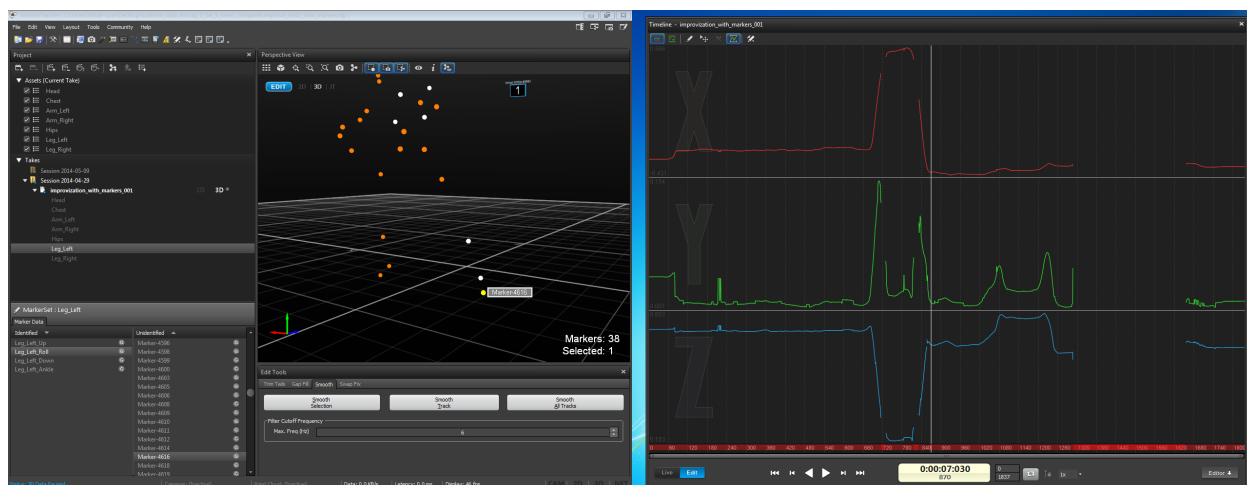
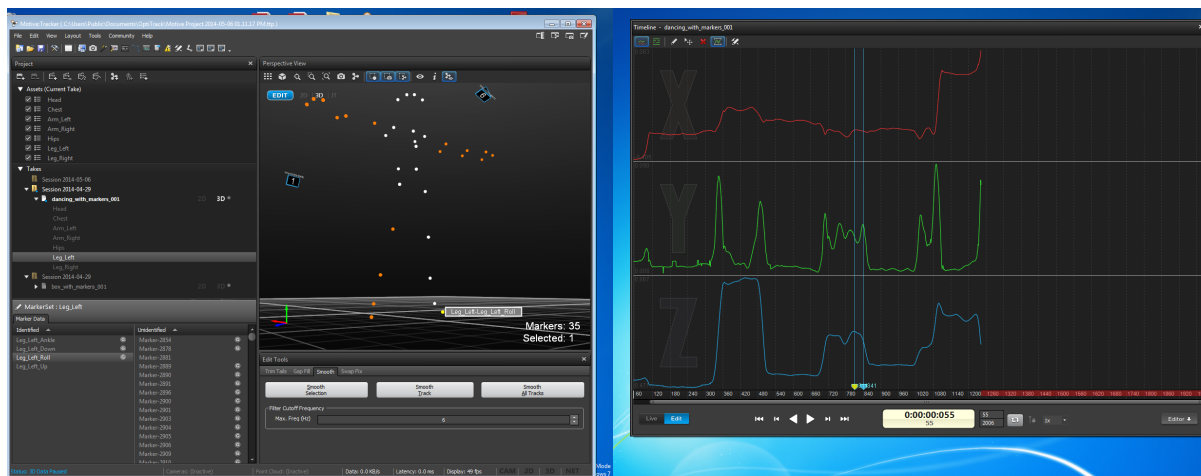
<http://www.naturalpoint.com/optitrack/static/img/calibrationSquareLarge.jpg>

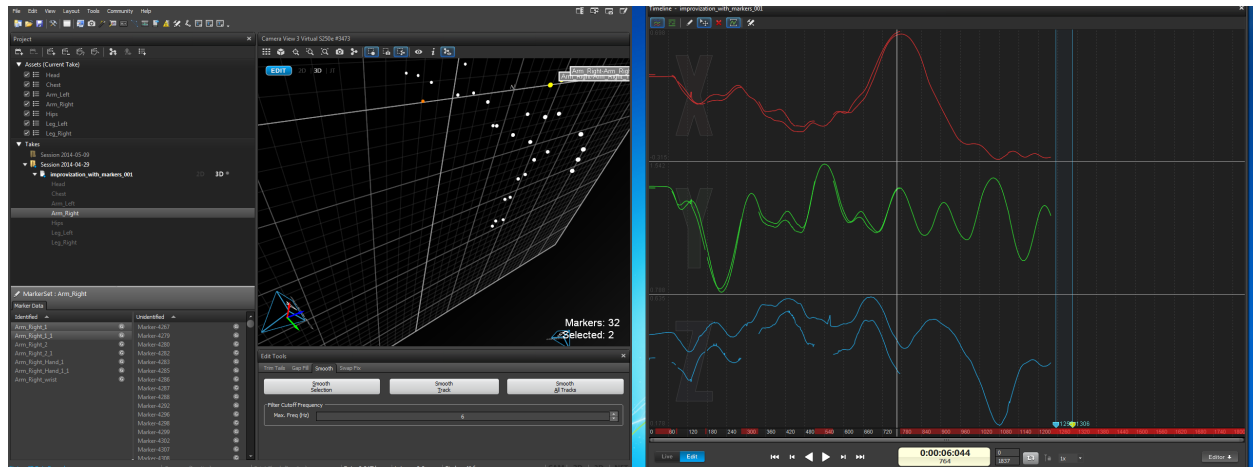
Motive:Tracker Calibration Layout example:



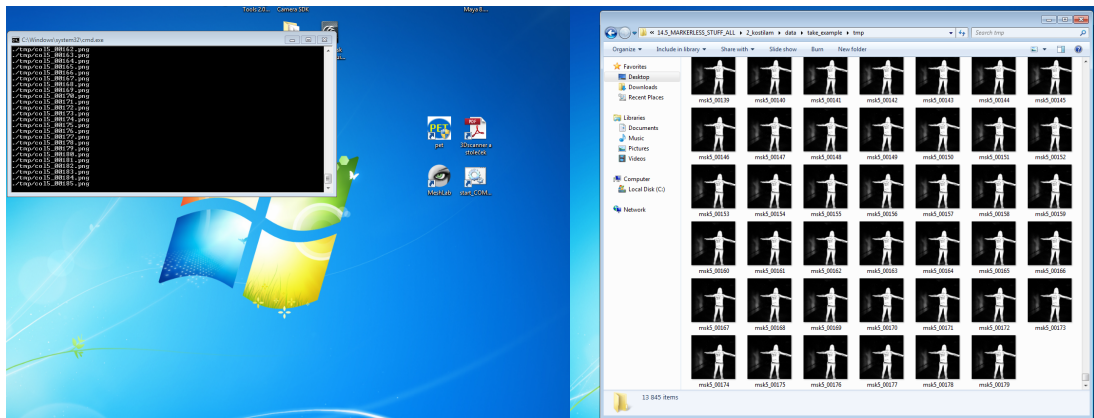
<https://encrypted-tbn0.gstatic.com/images?q=tbn:ANd9GcT3KbDatLTUhcAcY74cw8KXVpi6xMAX2QxtTcfZP6ZfZ0BZ9jiYNw>

Editing Phase, Session 2:

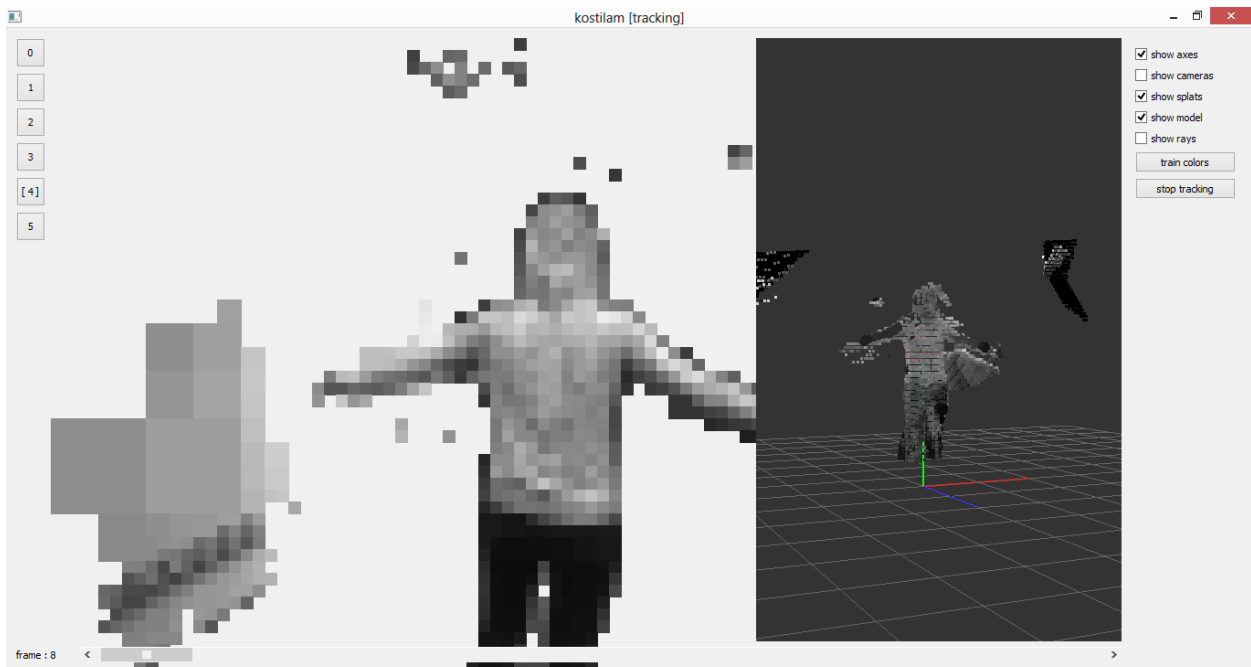
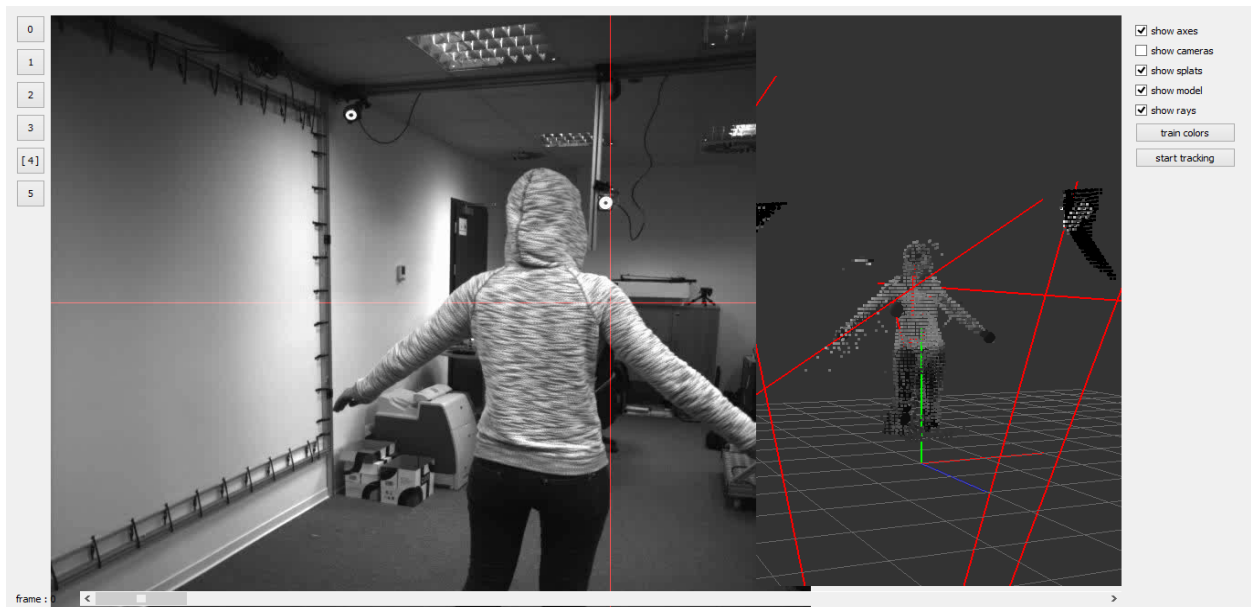


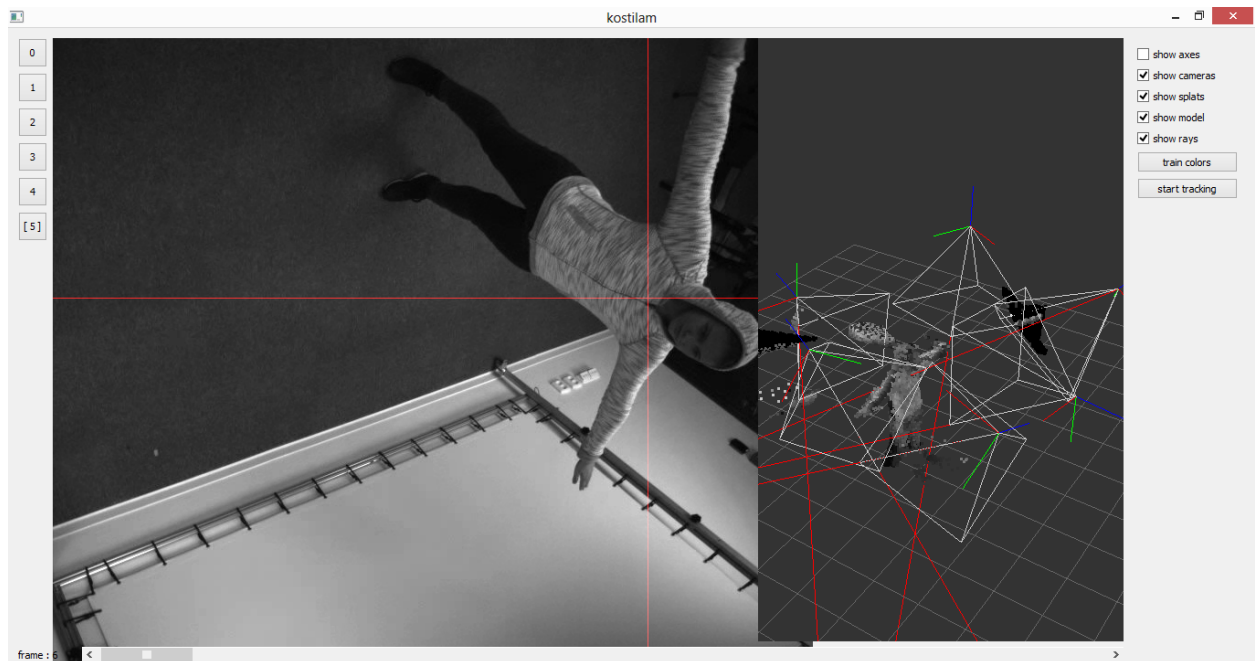
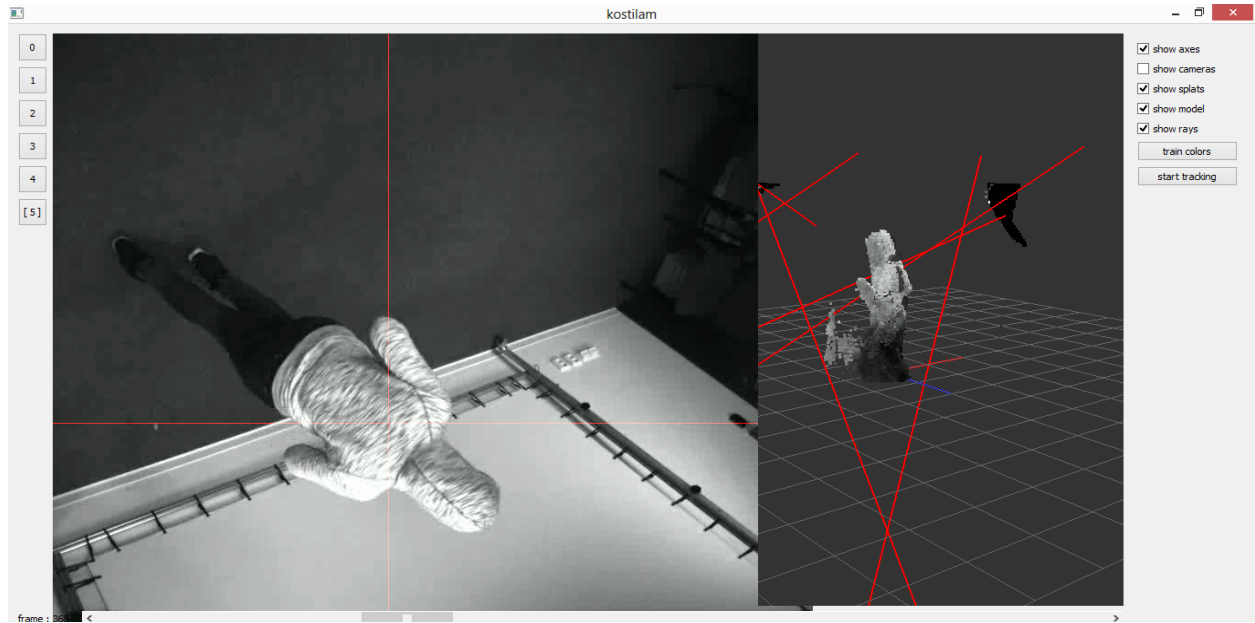


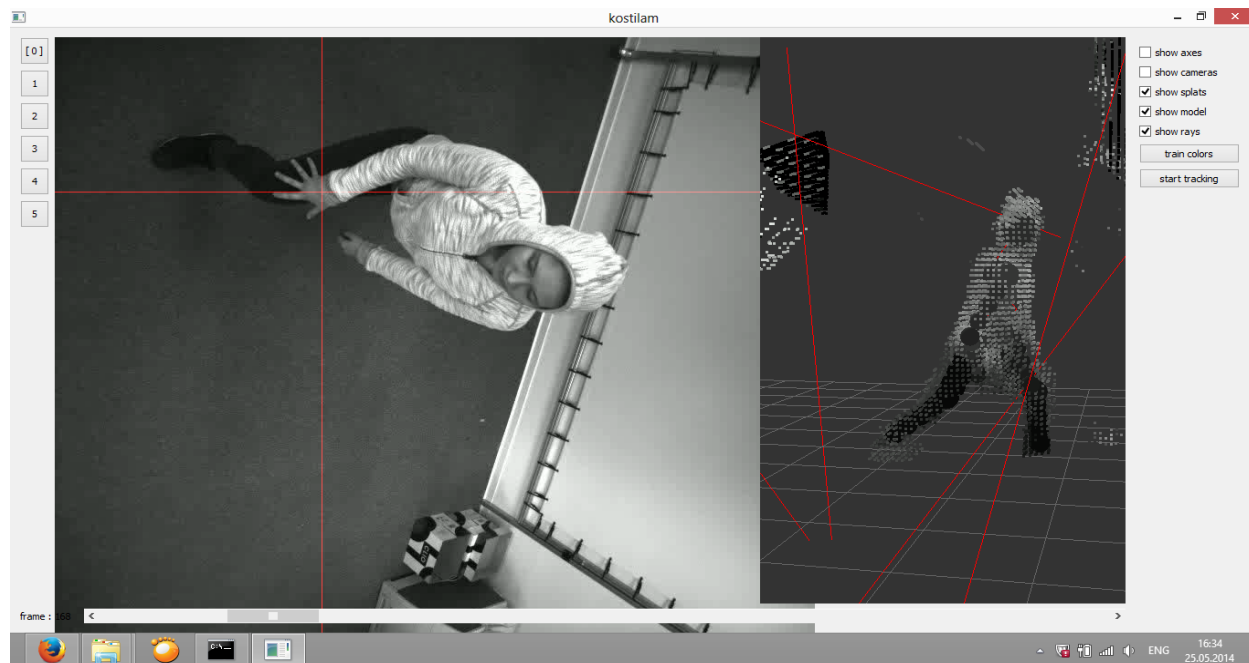
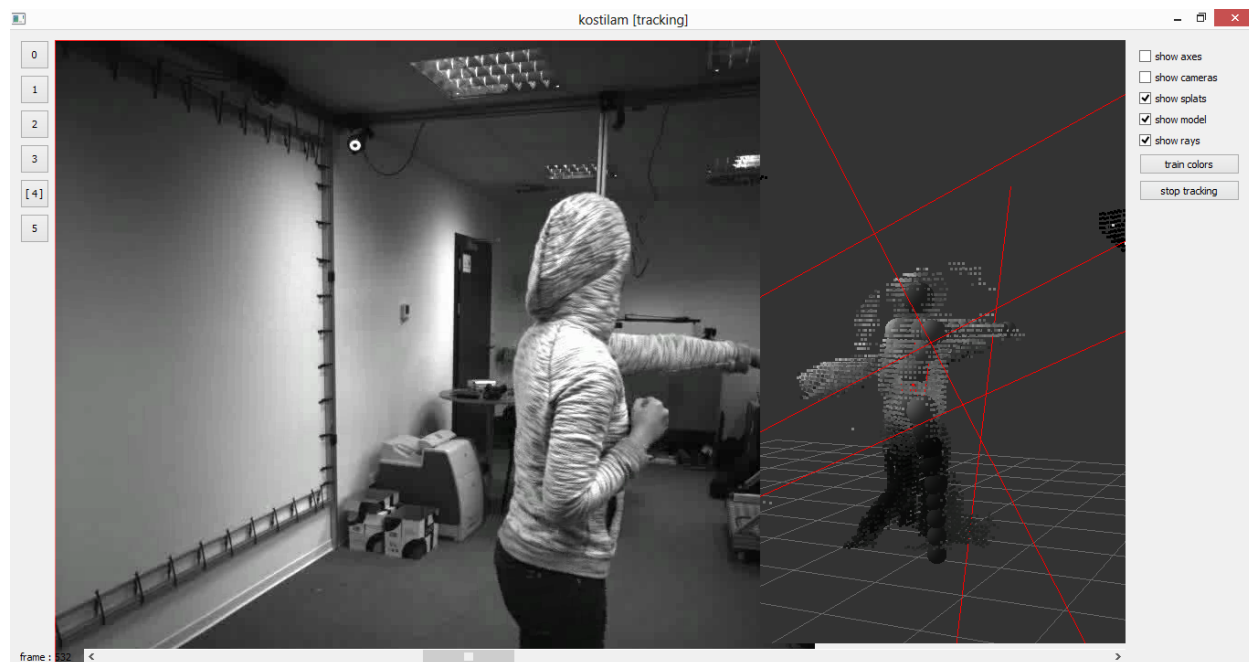
Markerless MoCap processing - masking:



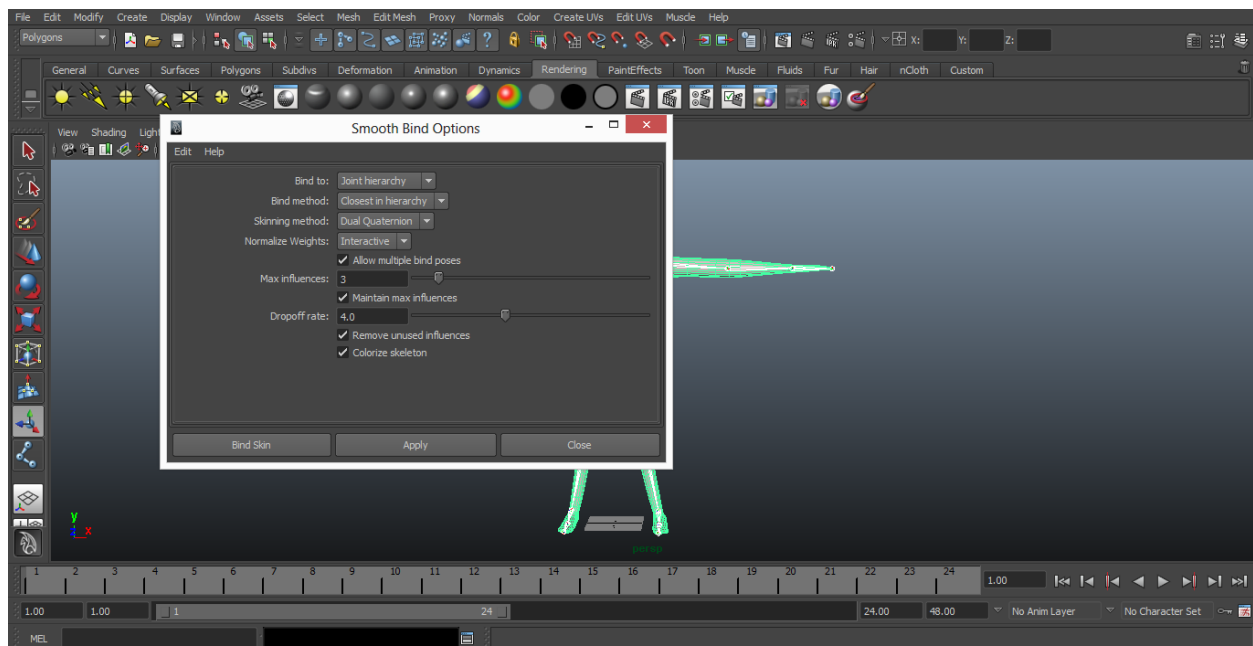
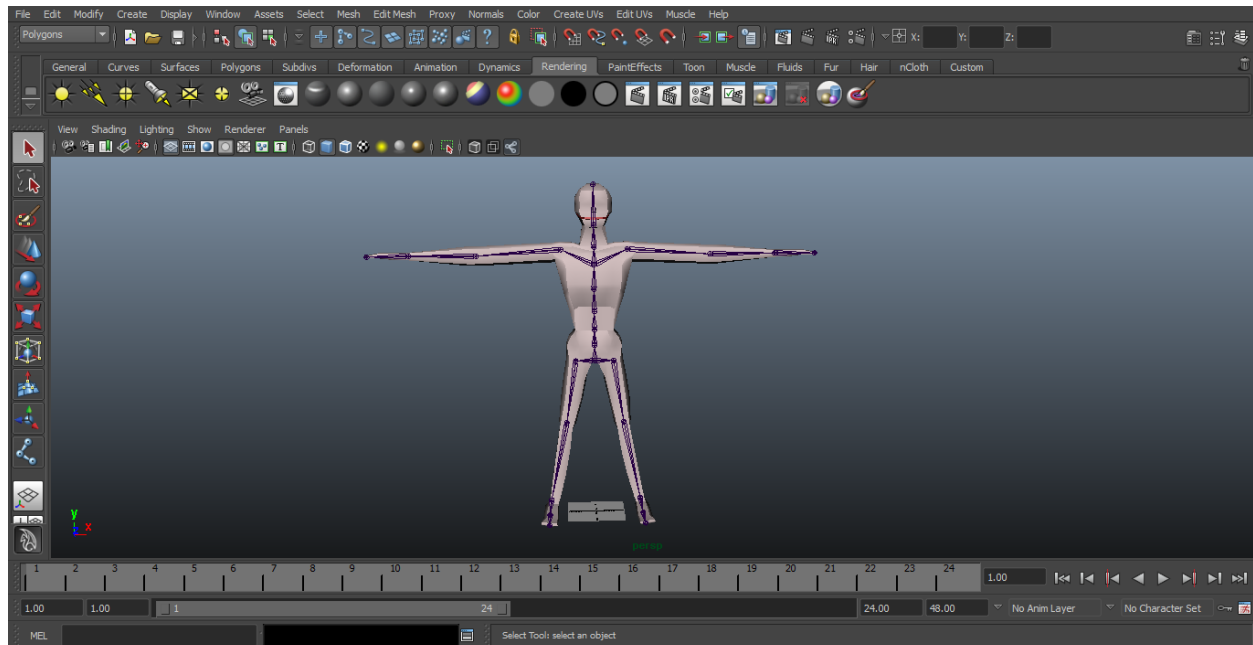
Markerless MoCap - tracking in Kostilam:



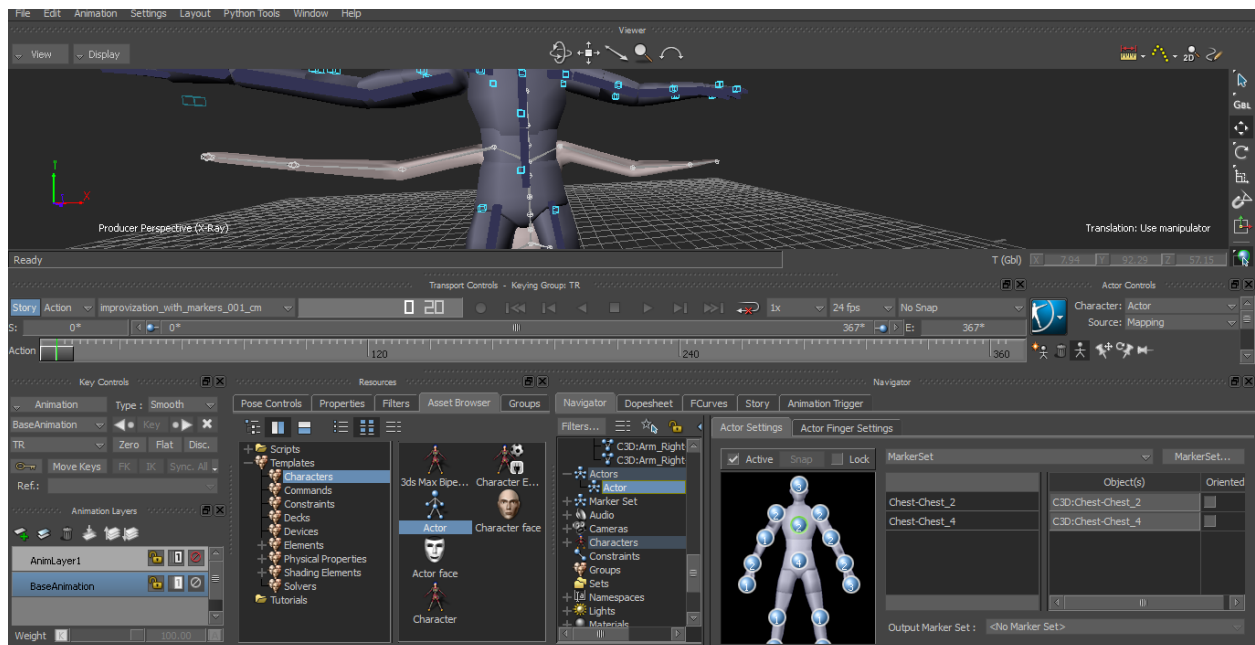
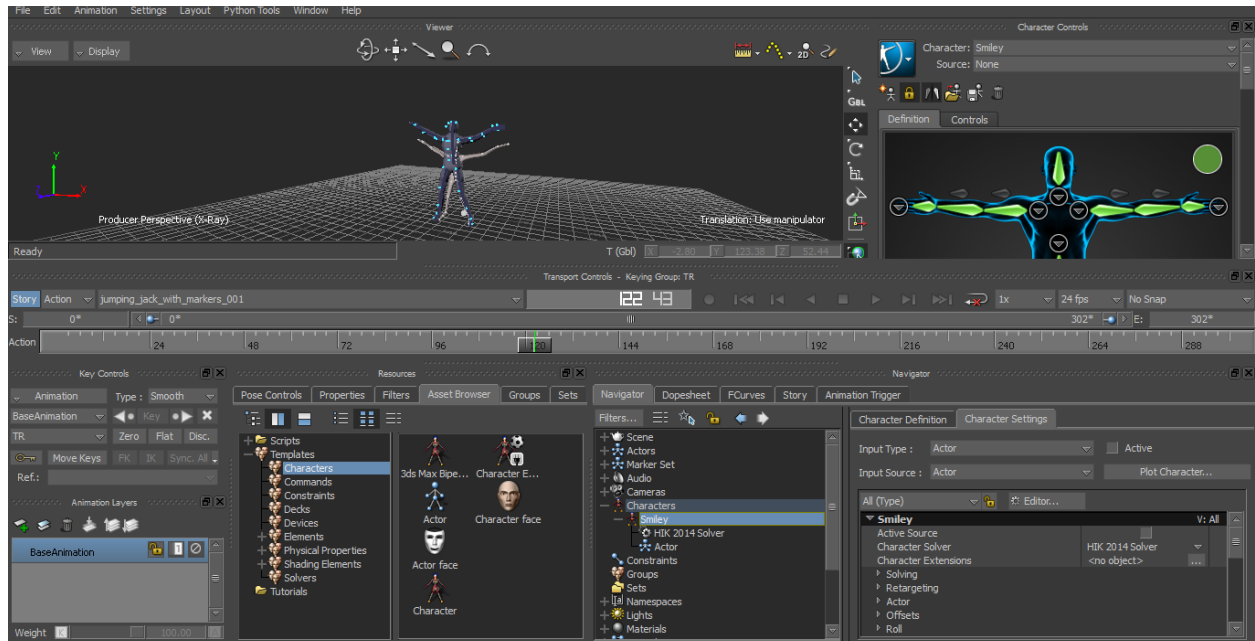




Rigging 'Smiley' version 2:



MotionBuilder data Application:



Disc Content:

model_Smiley_version_2: folder contains bc model Maya files;

Motive_Tracker_exported_with_markers: contains .c3d exported files from Session 1 and 2 - testing with markers;

Session 1: contains 5 folders for each motion sequence folder which contains:

- animated model as Maya ASCII file
- animated model as Maya binary file
- .avi rendered video
- MotionBuilder file: model with Actor as a source of animation;

Session 2: contains 5 folders with final sequences (testing with markers), same structure as Session 1;

Session 3: contains cameras.zip - Calibration layout for testing without markers for Kostilam, 5 folders with motion sequences, run-take_example batch file in each of the sequence folders starts generated Kostilam tracking;

videos: rendered animation sequences;

model_Smiley_version_1.zip: Maya model files of A7B39PRO version;

Motive_Tracker_src.zip: contains Session files related to MotiveTracker and backup.zip archive, containing copies of .ttp, TAKE files and .cal;