**Motion capture data processing and analysis**

Table of Contents

[List of Abbreviations 2](#_Toc520570227)

[1 Introduction 3](#_Toc520570228)

[2 Theory. Analysis. Aim. 4](#_Toc520570229)

[2.1 Motion Capture 4](#_Toc520570230)

[2.2 VICON 5](#_Toc520570231)

[Vicon spec // 8 camera optical based mocap, high fps, ... 5](#_Toc520570232)

[2.3 representation of skeleton by length and angles // bones or markers 5](#_Toc520570233)

[2.4 solving // process of translating the raw mocap data into CG character to create skeleton animation 5](#_Toc520570234)

[2.5 data format (.c3d) 5](#_Toc520570235)

[2.6 methods for segmentation and it is challenges // challenges with manual and automatic segmentation 5](#_Toc520570236)

[2.7 problems with acceleration computation // fuzziness 5](#_Toc520570237)

[2.8 methods for filtering 5](#_Toc520570238)

[2.9 methods used for describing the properties of each sign 5](#_Toc520570239)

[2.10 SVM 5](#_Toc520570240)

[3 Software Architecture 6](#_Toc520570241)

[4 Testing 7](#_Toc520570242)

[5 Conclusion 8](#_Toc520570243)

[6 Bibliography 9](#_Toc520570244)

# List of Abbreviations

3D Three-dimensional

MoCap Motion Capture

SL Sign language

fps frames per second

# Introduction

Motion capturing is a modern, fast developing data acquiring method capable to record movement in 3D. The detailed data that is retrieved from such recording is very useful not only for movie and game industry, but also in other fields as military, medicine and for validation and control of computer vision and robotics. The benefits of this method attract scientist to utilize it for linguistic analysis. Sign language is the primary alternative to a spoken language. Unfortunately, there are people for whom this is the only alternative that gives them the ability to communicate and share thoughts directly. Sign language uses manual movements and body language to communicate thoughts with others. The basic component of a sign language includes hand gestures, movements, orientation of fingers and hands, hand shapes and facial expressions to communicate certain feelings. Every region in the world has a unique spoken language and similarly, every region has a unique sign language. Thus, sign language varies from culture to culture and from region to region. People with speech and/or hearing impairment find it difficult to communicate with other individuals via sign language due to the inability of most of the people to understand sign language.

The purpose of this paper is to show the methods of studying and understanding the properties of signs from motion point of view and developing a tool for processing sign language data base. It is focused on hand movement and gestures analysis. It is cultural and regional independent, because it uses kinematic and statistical methods for processing the data.

Such database is set of dictionary (lexical items separated by default stance //not sure if it should be here or later in the main part) files with motion capture data. The tool will perform raw segmentation (in the first step) and fine segmentation (using acquired parameters from data) of dictionary items to extract the meaningful information for each sign which will be used for further analysis. It will show the challenges to determine the exact beginning and ending of the sign. The significant problem with the nature of the data – the containing of noise and methods for isolating that noise. Each sign will be processed for extracting its properties. Such as: if the sign is one handed or two, which is the dominant hand, hand location and orientation, finger orientation.

Later with the help of computer learning methods such as SVM the extracted information will be used to cluster signs for further processing.

# Theory. Analysis. Aim.

## Motion Capture

### History of motion capture

In general, the term motion capture (MoCap) is understood as the process of recording the movement of object, people or even animals. It is not specifically related to any device or approach. Today’s MoCap systems are product of many years of tinkering and innovation.

At the beginning MoCap analysis originate as gait analysis and animal locomotion around the year of 1872 and work of the photographer Eadwerd Muybridge. He used multiple cameras, triggered by strings to take pictures of moving bodies and animals. Capturing what human eye could not distinguish as a separate movement. His work “The horse in motion” was the first work recognized as a motion capture analysis. With a series of photographs of a galloping horse he proves that horses do have all four hooves of the ground during their running stride [[3](Muybridge's#_E._Muybridge,_)].

Since then the technology has developed to the form of devices designed for direct 3D recording. The biggest success and from where it gains its popularity is the usage in the movie industry. The first fully digital created character with the help of MoCap technique was Jar Jar Binks performed by Ahmed Best in Georges Lucas’ Star Wars Saga in 1990.

~~The usage of this techniques has expanded over the entertainment industry to sport industry, medicine and robotics. This leading to developing of different, more accurate and at the same time more affordable for wider use devices~~

Motion capturing started as biomechanics research tool, gained its popularity with its usage in entertainment industry and expanded into education, training, sports, and robotics.

There are different approaches to motion capture. Every system has its advantages and disadvantages. ~~Non-optical systems – Mechanical, Inertial and Magnetic. Optical systems are markerless, optical-passive and optical-active.~~ Optical passive system is considered as the most accurate and flexible technique and thus most common in the industry.

~~2.1.1 Markerless - This technique does not require markers to be worn and instead relies on software to track the subjects' movement. Varying tracking methods yield different results, but real-time and final data error ranges tend to be larger than marker-based solutions.~~

~~2.1.2 Inertial - This technique does not require cameras except as a localization tool. Inertial sensors are worn by the subject and the data from the sensors is transmitted wirelessly to a computer.~~

~~There are two variants of optical systems that use markers. The active optical system uses cameras that work in visible spectrum and markers in form of Light-Emitting Diodes (LED). The need of power source for each marker may be considered as disadvantage. Presence of wires may affect movements and their measurement. On the other hand, markers can be distinct easier during tracking, because LEDs use different wave length.~~

~~Usually cameras for MoCap systems work in infra-red spectrum and use markers coated with retroreflective material, also called passive markers. As disadvantage of this approach may be considered the need of external light at the scene. Often the external source of light is part of the camera’s body. Passive markers may be attached directly to the performer skin (in case of facial motion capture) or clothes. This method is considered as the most flexible and common type of MoCap.~~

## Optical-based motion capture system VICON

Motion capture technology was used to generate data files for both continuous and isolated utterances. The data used for this project was recorded at the University of West Bohemia with MX series device from VICON [4]. This is an overview of the basic principle of the technology on which the system is based. This system is based on optical- passive technology and was chosen for the sign language project because it suits the best for the purposes of the project.

The technology provides accurate data at fast sampling rates, and the same system can be used to capture the motion of a wide range of structures, including objects, animals, human bodies, fingers and faces. By using passive (reflective) markers, all processing is done externally, and the captured subject does not need to wear electrical equipment or wires. Which is advantage in capturing finger motion, because the presence of wires can impede the naturalness of movements. The system is comprised of eight specialized infra-red cameras along with computers and software for image analysis and processing. The cameras detect small markers placed on strategic locations on the captured subjects. In this case passive markers are used, they are coated with retroreflective material, and this requires the cameras to emit the light, which is reflected back and detected. The external source of light is part of the camera’s body.

The detected information is then processed in dedicated hardware using software provided by the manufacturer to triangulate the 3D locations of the markers. The exact process and algorithms are know-how of devices’ manufacturer, but it is based on stereoscopic vision.

After reconstructing the 3D points into a point cloud, the system needs to determine which point is which and label each point with a marker id. This process is commonly referred to as marker labelling. The marker data is then used to estimate the kinematic motion of a model of a human skeleton. The estimation in form of bone lengths and joint angles can be used for further processing or animation. This process is called solving. The mentioned software is able to provide semi-automatic calibration of human subject for general movement.

Although this technology comes with highly accurate results and it is very flexible there are some disadvantages which may result in poor data quality or extensive costs in manual post-processing. As every vision-based technology it needs clear line of sight and occlusions may cause serious challenges. Markers placed on fingers are especially problematic and often suffer from self-occlusions when the fingers are bent or the hands are facing towards the body or with palm-up [[5](#_N._Wheatland,_Y.)]. Occlusions do not only cause problems with missing data, they also make the labeling process more difficult as this reduces the available information for inference. Further challenges arise in situations when several markers come in close contact (such as clapping hands).

To summarize, optical motion capture provides highly accurate data, but may require a large amount of manual post-processing

## sign language avatars

## Sign language analysis // of segmentation problem // nature of signs

Sign language is used by millions of people around the world. It is used to facilitate communications with people with speech or hearing impairments. There are different sing languages as there are different spoken languages. Also there is common misconception that sign language is dependent on spoken language, that it is spoken language expressed in signs [[8](#_Perlmutter,_David_M.,)]. Linguists has studied and proved that sign languages exhibit the fundamental properties that exist in all languages and there are similarities between both forms, but there are some basic differences. The linguistic mechanism in both is different and therefore it causes difficulties for people with such impairments, especially for those who are born this way, to use even the written form of a spoken language [[9](#_Sandler,_Wendy;_&)]. In spoken language units are organized sequentially (it is not possible to say two different words at the same time), but in sign language the meaning of one unit may be carried by the shape of the hands and their position or/and by the position and movement of head and mimics. These two components can be carried simultaneously.

All these circumstances not only put a communication barrier between people using sign language and majority of hearing community, but also restrict them from most sources of information. This led scientist to develop writing system to represent signs. Such system is the Hamburg Notation System (HamNoSys). It is a work of scientists from University of Hamburg. It is an alphabetic system that describes signs on a mostly phonetic level. It is based on Stokoe’s notation system [[1](#_T._Hanke,_“Hamnosys-representing)]. It is designed to be usable in variety of context with the following goals in mind:

* International use - HamNoSys transcriptions to be possible for virtually all sign languages in the world.
* Iconicity – the large number of parameters variations did not allow the use of standard alphabet, newly created glyphs were created in the way that helps to memorize or deduct the meaning
* Economy – notation of signs should make use of principles as symmetry conditions, thus resulting in shorter notions.
* Integration with standard computer tools – It should be usable computer –supported transcription, standard text processing and database applications.
* Formal syntax – The notation language should have well-defined syntax, its semantics should follow the compositionality principle.
* Extensibility – As SLs are developing and differ from each other, HamNoSys should allow both for a general evolution and specializations. [[1](#_T._Hanke,_“Hamnosys-representing)]

Because of all described goals of HamNoSys it is liked and preferred from scientific community working in the domain of language analysis and synthesis, although it is not very accepted of deaf community.

I used this notation system as guidelines for describing the properties of signs.

## methods for segmentation // kinematic movement analysis

## filtering of acceleration

# Software Architecture

## data format (.c3d)

## methods for segmentation and it is challenges // challenges with manual and automatic segmentation

## problems with acceleration computation // fuzziness

## methods for filtering

## methods used for describing the properties of each sign

## SVM

# Testing

# Conclusion

# Bibliography

#### T. Hanke, “Hamnosys-representing sign language data in language resources and language processing contexts," in LREC, vol. 4, pp. 1-6, 2004.

#### J. McDonald, R. Wolfe, R. B. Wilbur, R. Moncrief, E. Malaia, S. Fujimoto, S. Baowidan, and J. Stec, “A new tool to facilitate prosodic analysis of motion capture data and a data- driven technique for the improvement of avatar motion," 2016.

#### E. Muybridge, “Muybridge's Complete Human and Animal Locomotion”. Collections of ne art in Dover books, Dover Publications, 1979

#### “Go Further with Vicon MX T-Series” Vicon Motion System Limited, rev. 1.3 August, 2010

#### N. Wheatland, Y. Wang, H. Song, M. Neff, V. Zordan, and S. Jörg, “State of the art in hand and finger modeling and animation” Computer Graphics Forum, vol. 34, no. 2, pp. 735–760, 2015

#### M. Dilsizian, Z. Tang, D. Metaxas, M. Huenerfauth, and C. Neidle, “The importance of 3D motion trajectories for computer-based sign recognition" 7th Workshop on the Representation and Processing of Sign Languages: Corpus Mining, The 10th International Conference on Language Resources and Evaluation (LREC 2016), 2016.

#### T. Hanke, S. Matthes, A. Regen, and S. Worseck, “Where does a sign start and end? segmentation of continuous signing" 5th Workshop on the Representation and Processing of Sign Languages: Interactions between Corpus and Lexicon, Language Resources and Evaluation (LREC 2012), 05 2012.

#### Perlmutter, David M., "What is Sign Language?"LSA, 12 April 2014. Retrieved 4 November 2013.

#### Sandler, Wendy; & Lillo-Martin, Diane. ,“Sign Language and Linguistic Universals” , 2006. Cambridge: Cambridge University Press.