

Pose and motion capture technologies

Arun Fabian Panaite^{1,*}, *Sebastian Rosca*¹ and *Remus Sibişanu*¹

¹University of Petroşani, Department of Automation and computers, 20 Strada Universităţii, Petroşani, România

Abstract. The multitude of sensors, transducers and data acquisition systems used in the ever-increasing domains of today's industry, were designed and created on a purpose-based agenda. Most of these technologies were rarely used for different purposes than the ones they were meant for. This research, explores the possibility of using those technologies for the purpose of motion capture. The results of the motion capture data acquisition can further be used for other purposes, itself. Because the motion capture data isn't calculated, but only captured, the compression algorithms may result in mathematical formulas of motion, which can be the same formulas of calculating the motion synthetically. The same way the parameters and formulas are being calculated for a robot's movement, motion can be captured, and the key position parameters can be selected, and implemented on the movement, while not only making the movement more natural, but saves time during the programming process.

Keywords: health, safety, computer vision, wearable, sensors

1 Introduction

Ever since motion capture became an idea, scientists as well as other minds began to materialize this idea using various techniques and approaches that used the latest available technology. In the past decade, motion capture began to be used in multiple domains, such as sports, entertainment and medicine. Wearable devices for real time movement analysis and feedback, although with the limitation of questionable accuracy of measuring the short, high intensity movements [1] can be used, while incorporating MEMS (Micro Electro Mechanical Sensors) to have the tri-axial movements and positions recorded and analyzed.

Since there's yet no set of rules that prevent or direct the procedures and data interpretations that the users make, it is the user's knowledge and experience to dictate whether the data is to be trusted, having the unalienable right to choose to allow the technology to obtain the data. MoCap (short for Motion Capture) is represented by the entirety of measurements and digitally tracking and recordings of moving objects, bodies and living things in a three-dimensional environment (space) [2]. For a more accurate set of

*Corresponding author: fabianpanaite@upet.ro

results, newer technologies employ hybrid systems, built around daedalian cores, such as machine learning, Kalman Filters, mazy hierarchical clustering and many more.

By domain, MoCap technology can be used in healthcare, for diagnosis and treatment of physical infirmity, reviewing and aiding the restoration of the motor functions of patients; in sports, to analyze and enhance athlete dynamism and posture; in entertainment, gaming projects and businesses, for the purpose of bringing more viridity to character motion; in robotics, automation, construction and automotive applications to optimize and naturalize the motion of machines, considering that nature is trying very hard to make us succeed [3].

Microsoft Kinect, the gesture sensitive controller was developed in 2010 using an RGB-D camera. Lately, the controller started being used for more than its original purpose, thus, the controller made for entertainment and gaming, started being used for motion capture howbeit it wasn't expressly designed to do so, not only MoCap being just one of the collection of plans it was used for, 3D scanning, stroke recovery, sign language communication, gesture based security, virtual reality interaction, virtual surface touchscreen applications and many more are already developed, as an aperture for its use in a diversity of domains, [4] including the proliferating towards inexorable hobbies of "lifelogging" (a term used for the activities in which people record their daily routines) [5] thus, also entering the domains of safety and security.

The behavioral examination of certain socio-demographic groups and their mobility patterns, allow to better determine or predict the capabilities and responses for hazard avoidance. Pose measurement is an essential part of the MoCap process, yet pose measurement can be used in other applications, in the domains of robot control, or aircraft flight control, accomplished by detecting changes between images and transforming them into geometry data [6].

In the field of pose measurement many physical parameters can be detected, measured and processed, using the right sensors, in addition to the magnetic, ultrasonic, mechanical and optical ones, depending on the environmental perils and the jamming resistance of the sensors. MoCap can also be used in different stages of the domains that involve the need of synchronous movement [7]. Not only to train the user's behavior against the exit from synchrony, but also to enhance the beginning and the re-entry into synchrony not only by the provided feedback, but by measuring the times and the surges of progress and advancements towards higher synchrony rates. The relationship between the mind and body can also be analyzed, [8] by assessing movement and recording, observing and profiling the minds and bodies of the recorded people.

Detecting the correlations between the postures and the personality types, is useful for observing and managing the "self-image" [9] as it has the role of amplifying the person's self-awareness. There are situations in which the motion capture data becomes corrupted, unreadable, or incomprehensible. [10] Camera occlusions, sensor connection errors, or skeletal position misinterpretations and even system software, hardware or file management and storage errors are the main sources of MoCap data corruption. With cognizance to the possibilities of direct intervention to the MoCap data, while excluding the indirect data sources, means of data recovery were predicted, using skeletal constraints. MoCap and Motion Analysis cannot be limited to the ground and to the air. As proof, cameras were used under water to perform 3D Motion Analysis, the accuracy being comparable with the traditional motion capture systems.

Video-capture based technology proving itself to be useful in watersports domain for gesture study and athletic performance evaluation [11]. Marker-less motion tracking [12] is simultaneously a MoCap method, and a MoCap data correction method as an application for MRI (magnetic resonance imaging). Eliminating markers from the MoCap process is not an easy task, but it has recently become a requirement, since marker misplacement [13] alters the data.

2 State of the art

In the first reference's study, a total of 19 accelerometer sensors were mounted on a moving table, and moved on all three orthogonal axes, revealing that their intra and inter-device precision and reliability, since they were wearable devices, was easily maintained in both laboratory conditions and outdoor conditions, adding the fact that firmware updates enhanced not only their communication, but also their precision.

The second reference study suggests that depending on the purpose, the MoCap is mainly (mainstream) done using one of these two technologies:

- a. IMU (inertial measurement units);
- b. Camera-based MoCap.

The IMU MoCap technology is mainly used in the Entertainment and Gaming industry, and in the domain of various Sports, while the camera-based MoCap is frequently used in robotics control. A study involving the Microsoft Kinect had demonstrated on multiple occasions that albeit the Kinect is not intended for very precise MoCap, but only gesture detection, its sensors and cameras are incredibly capable of very precise motion capture, as seen in the figure below.

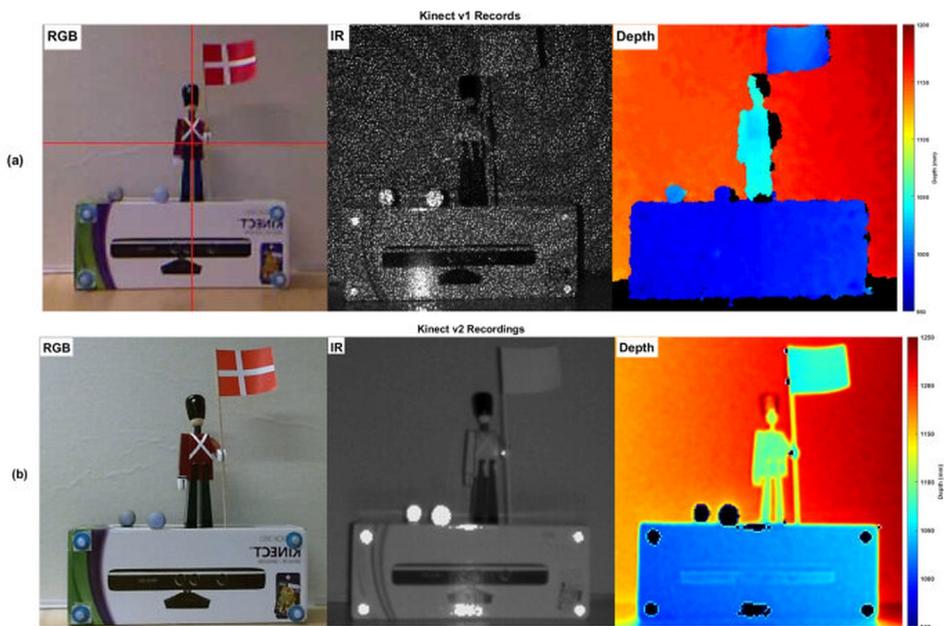


Fig 1. Microsoft Kinect records. (a) RGB and infrared records of Kinect v1 and corresponding depth map. The black pixels in the depth image reflect areas of unknown depth, which in this record is due to the out-of-record boundary, noise, and shadow effect on the left side. (b) RGB and IR images of Kinect v2 sensor and the corresponding estimated depth information (right panel), the black pixels in the depth image reflect areas of unknown depth.

In a multi-sensor movement analysis for transport, safety and health applications the scientists have designed a problem-solving framework for the indoor and outdoor movement detection reading, with the images captured by a lifelogging wearable device as the one seen in the figure 2.

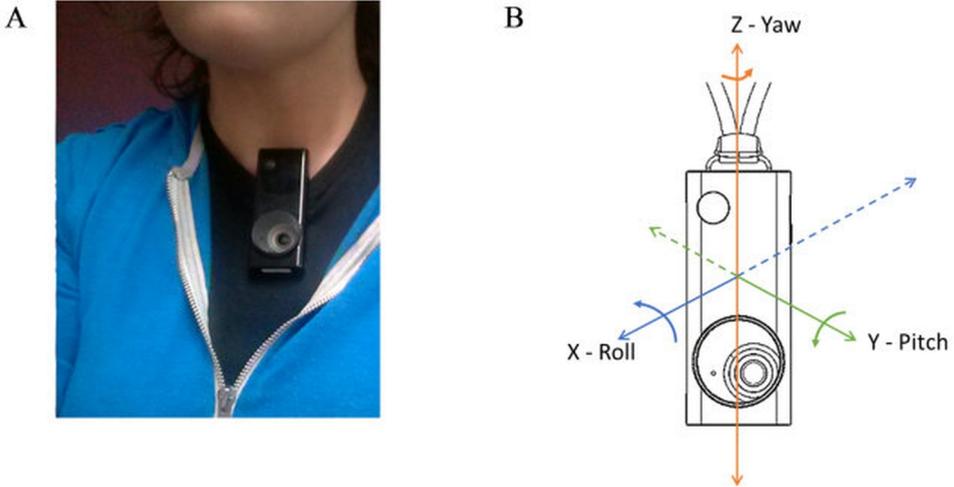


Fig 2. A. Autographer—lifelogging device. B Orientation determined by Autographer manufacturer using magnetic field and acceleration where Yaw is an Azimuth (0–360 degrees), Roll (-90-90 degrees) and Pitch (-180-180 degrees).

Global Control Point Calibration used in Pose Measurement can easily solve the pose measurement errors and it does it by adding spatial layers to the MoCap, in a way that a certain distance is at a certain layer. Regarding the Exit from Synchrony in Joint Improvised Motion, the scientists have presented a mathematical mechanism that captures aspects of exit-from-synchrony in human motion, as it has been studied, and proven that humans enter and exit synchrony repeatedly, involuntarily, rather than voluntarily and maintaining synchrony. In the research of a Camera-Based Monitoring of Bodily Movements using Motion Silhouettes the scientists have validated the Motion Tracker software, to be the simple, highly very effective software program, that utilizes established CVT (computer vision techniques) for pose estimation and motion of a person in a video. This system works with any camera, and existing videos, without using intrusive or expensive devices.

The robust and automatic motion-capture data recovery using soft skeleton constraints a model averaging method that shows the overall approach, which can be split into two main steps, as seen in the figure 3.

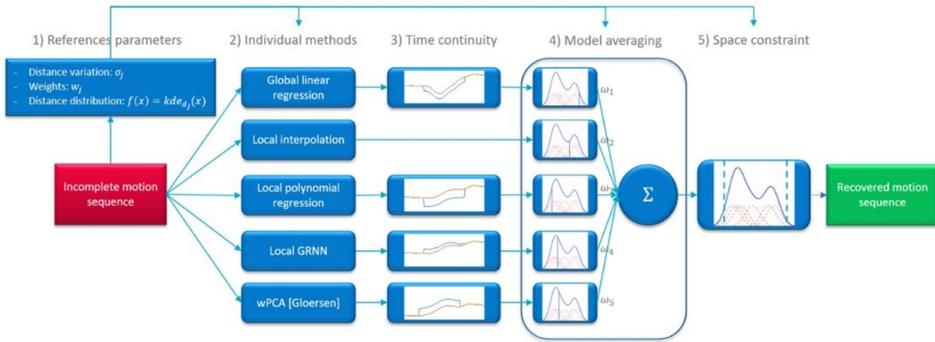


Fig 3. Block diagram of the proposed method.

The overall process can be divided in five steps:

- 1) Extraction of marker trajectories parameters.
- 2) Individual recovery models.
- 3) Time constraint: trajectory continuity.
- 4) Distance-probability weighted averaging.
- 5) Spacing constraint: reference marker distance likelihoods.

The research of Action Sport Cameras as an Instrument to perform a 3D Underwater Motion Analysis is using GoPros mounted and calibrated under water to capture images and videos, controlled by Wi-Fi remotes, and saving the videos as AVI format.

The more often calibrations were made, the more optimal were the results. Marker-less motion tracking and correction for PET, MRI, and simultaneous PET/MRI was a demonstration and evaluation of the first marker-less motion tracker compatible with PET, MRI and simultaneous PET/MRI for motion correction of brain imaging which used methods such as the Multiple Acquisition Frame method and Real-time Motion estimation. In the paper about the possibility of fusing motion capture and 3D medical imaging, reducing the extrinsic variability due to marker misplacements, the approach was consisting in using anatomical landmarks identified with 3D medical imaging to correct marker misplacements as seen in the figure below.

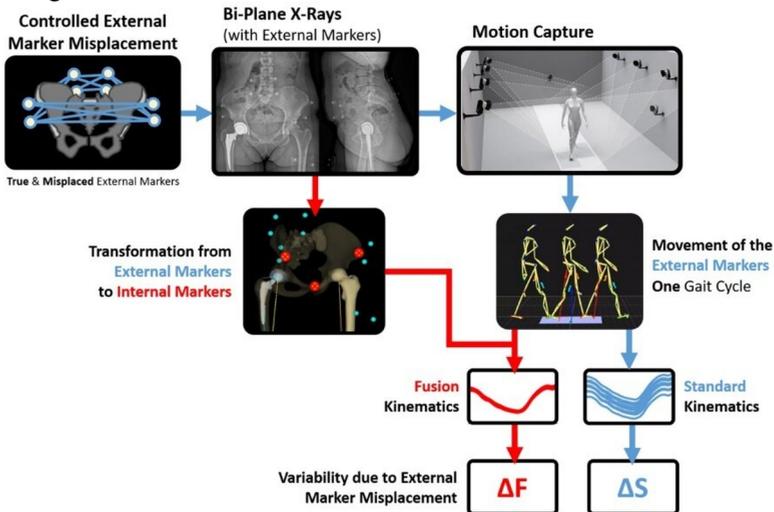


Fig 4. Workflow of the standard method and fusion method.

3 Methodology

Paper [1] has nineteen portable accelerometers analyzed, and each was identically and rigidly mounted onto aluminum brackets that were bolted to an electrodynamic shaker table, the same model and design that NASA is using for testing their microsatellites. A reference accelerometer was used, to allow the scientists to differentiate the precision of the other nineteen, the devices were subjected to various oscillations in each of the three orthogonal axes.

In most of the other reviewed studies, pose and position estimation was analyzed using various techniques, which all had mostly the same bold, direct question at the base: Which of these techniques provides the most accurate, yet accessible MoCap data? The answers varied, depending on the method as most of them have advantages and disadvantages, and the differences in the methods were not very high.

Most of the techniques had employed either IMUs, Cameras, or certain combinations of IMUs, cameras, and IR technology. In approximately 30% of them, the purpose of using the specific technology was completely justified and in the other approximately 70%, using the technology as a first try, although not always mentioned had palpable results.

4 Mainstream vs Alternative perspective comparison

Starting from the premise that the most generally used sensors in most domains of industry were IMUs, and marker-based sensors, some of the marker-less sensors and camera systems were a distinctively impressive option, to say the least. The main specifications of the MoCap approaches are corelated in the table below, with short explanations at the end, for the approaches that bring fog to interpretation.

Table 1. Most used MoCap systems

Key Features	Sensors		
	IMUs	Camera -Based	
		Marker-Based	Marker-less
Accuracy	High (0.75 ^o to 1.5 ^o) ³	Very high (0.1 mm and 0.5 ^o) ¹ ; subject to number/location of cameras	Low (static, 0.0348m) subject to distance from camera
Set Up	Straightforward; subject to number of IMUs	Requires time-consuming and frequent calibrations	Usually requires checkerboard calibrations
Capture Volumes	Only subject to distance from station (if required)	Varies; up to 15 × 15 × 6 m ¹	Field of view: 70.6 ^o × 60 ^o ; 8 m depth range ⁵
Installation Cost	From USD 50 per unit to over USD 12,000 for a full-body suit ⁴	Varies; from USD 5000 ² to USD 150,000 ¹	USD 200 ⁵ per unit
Ease of use and data processing	Usually, raw sensor data to ASCII files	Usually highly automated, outputs full 3D kinematics	Requires custom-made processing algorithms

Invasiveness (body)	Minimal	High (markers' attachment)	Minimal
Invasiveness (workplace)	Minimal	High (typically, 6 to 12 camera systems)	Medium (typically, 1 to 4 camera systems)
Line-of-sight necessity	No	Yes	Yes
Portability	Yes	Limited	Yes
Range	Usually up to 20 m from station ³ (if wireless)	Up to 30 m camera-to-marker ¹	Low: skeleton tracking range of 0.5 m to 4.5 m ⁵
Sampling rate	Usually from 60 to 120 Hz ³ (if wireless)	Usually up to 250 Hz ¹ (subject to resolution)	Varies; 15–30 Hz ⁵ or higher for high-speed cameras
Software	Usually requires bespoke or off-the-shelf software	Requires off-the-shelf software	Requires bespoke software, off-the-shelf solutions not available
Noise sources and environmental interference	Ferromagnetic disturbances, temperature changes	Bright light and vibrations	R-interference with overlapping coverage, angle of observed surface
Other limitations	Drift, battery life, no direct position tracking	Camera obstructions	Camera obstructions, difficulties tracking bright or dark objects
Favored applications	Activity recognition, identification of hazardous events/poses	Human–robot collaboration, robot trajectory planning	Activity tracking, gesture or pose classification
¹ Based on a sample layout with 24 Prime x 41 Opitrack cameras. ² Based on a sample layout with 4 Flex 3 Opitrack cameras. ³ Based on the specs of the Xsens MTW Awinda. ⁴ Based on the Xsens MVN. ⁵ Based on the Kinect V2.			

5 Results and Discussion

The results in the studied researches are that the Catapult OptimEye S5 accelerometer units provided exemplary reliability [1]. The data acquisition was a reliable process, and the acquired data was more than acceptable to work with, as seen in the accuracy table.

Table 2. Accuracy of wearable devices compared to the reference accelerometer (percent difference).

	0.1g	0.5g	1.0g	3.0g
X-direction	22.3%	4.1%	2.6%	2.4%
Y-direction	23.5%	4.8%	2.6%	1.8%
Z-direction	4.9%	1.9%	1.0%	1.0%

Health and safety in the working environment, had been discovered to be the most facilitating area for research. In clinical environments, the use of sensors had become almost unescapable, but the use of multiple sensors to identify just one parameter, to measure and

confirm the measurement was rarely an option, excluding military measurements, or the domain of aerospace, after all, the health and safety of workers is just as important.

The study [4] analyzed the impact of a marker-based motion capture system, based on the Microsoft Kinect v2 sensor in five consecutive distances. It has been observed that the Qualisys cameras may have a negative impact on the Kinect v2 measurements and recordings and post-processing calculations and as a consequence, certain degrees of uncertainty may be imposed on the measurements of the Kinect 2.

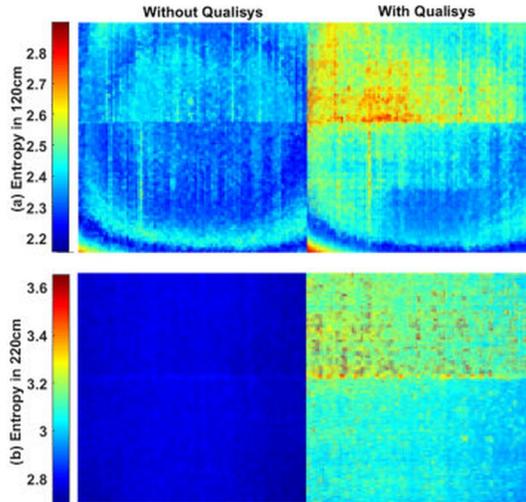


Fig 5. Comparing the impact of Qualisys on the entropy of Microsoft Kinect depth records. Where Kinect was placed at (a) 120cm, (b) 220cm,

The indoor and outdoor activity detection in [5] showed that even using dispersed sensor data can actually provide high precision results. In the classification using multi-sensor readings have equivalent results to previous studies that used only acceleration even with much higher reading intervals. In consequence, it has been observed that the people who were active outdoors were indeed active indoors as well.

For data acquisition an infrared camera [6] was used, with an 80° FOV (field of view) with its parameters calibrated as in the next table.

Table 3. Internal camera parameters

Focal length(mm)	8.283	Distortion factor k1	1.442e-003
Image center Cx(pixel)	419.47	Distortion factor k2	-4.567e-005
Image center Cy(pixel)	246.43	Distortion factor s1	1.365e-004
Non-perpendicularity factor	1.039	Distortion factor s2	-1.902e-004

The infrared LEDs are feature points for positioning. The phenomenon [7] of exit-from synchrony was focused on the study, which is dedicated to understanding how, in motion-coordination tasks, people exit and re-enter periods of high zero-phase synchrony.

A mathematical model of the synchrony exit and re-entry was developed to determine the cause and add a predictor + corrector sum to not only identify the cause, but also to correct the effect. As an explanation to the mirror game-based behavior, the model needed a predictor-corrector based design. In the model, the rate of change of velocity of the tracker

v_1 is provided by a corrector term $f_1(t)$ and a predictor based on a sum of periodic functions with time dependent amplitudes $A_1 n(t)$

$$\frac{dv_1}{dt} = \text{predictor} + \text{corrector} = \sum_n A_1 n \omega \cos(\omega n t) + f_1(t) \quad (1)$$

There is, however a velocity error between the tracker v_1 and the stimulus v_2 , which are integrated by the corrector:

$$\frac{df_1}{dt} = k_1(v_2 - v_1) \quad (2)$$

This approach [7] can be used for motion modelling and simulating in treatment and rehabilitation computer set-ups that use the mirror game. Experiences in the mirror game suggest that by the present mathematical models there may be additional discretions regarding the exits-from-synchrony, that cannot be captured yet. The movement's measure [8] obtained from the Motion Tracker program, has strong links to a proprietary BPMS (body pressure measurement system) and the Kinect's head-tracking. This solution may be very useful to research in low budget laboratories, where the burdens of all-day-long programming are reduced, operating a simple plug-and play system, while also allowing the designing of simple and ecologically valid research. It's inexpensive as a result of no special hardware requirements, other than a web camera. It is non-intrusive and non-invasive since it doesn't require any wearable devices. It has become a normality for it to be portable and easily deployable, in comparison with some alternative technologies.

The Motion Tracker software has its limitations. The software is only compatible with a certain operating system, and it was written in an old programming language which makes it vulnerable. This measure of movement requires having videos of a certain quality and resolution; therefore, it is imperative that noisy backgrounds (with poor quality image) have to be reduced. The PMA (probabilistic model averaging) [10] method provides several advantages compared to the available conventional state of the art by being fully automatic and reducing the need for pre-trained models or a knowledge base while permitting the use on any marker set MoCap data. Furthermore, in comparison to distance-based reference markers, the PMA, partly uses the skeletal constraint rules.

As a limitation, the most fit parameters could depend on the MoCap data, and metadata, like the number of markers, their particular placement, the motion's speed and complexity, data accuracy, marker vibration noise, or camera noise and quality. As an example, a wider brink for the linear regression model could be attempted for MoCap data with fewer markers, and the optimal smoothness parameter for LGRNN (Local Generalized Regression Neural Network) could indeed be dependent on the motion's smoothness. Performing a 3D kinematic [11] analysis in sports, for instance, swimming, requires high accuracy of reconstruction. The use of nonlinear camera calibration improves the accuracy results found in laboratory conditions. When using optoelectronic systems and industrial cameras, small reconstruction errors (0.58 mm to 1 mm) have been obtained. The article was made as a demonstration of the feasibility of the quantitative 3D measurements performed underwater, using cameras. With a convenient calibration methodology, the cameras can become a very accurate metric system. In comparison with optoelectronic devices, designed for 3D motion analysis, the sports camera-based technique is appropriately, cost effective, reduced in size, highly portable, wireless and in this case waterproof. The potential applications for motion analysis can be a water sport.

Performing real-time motion tracking and monitoring compatible with PET and MRI scanners. [12] Optimizing the design makes the motion tracker beneficial for the MRI environment while not compromising PET sensitivity. It was the first time that a motion

tracking system was successfully demonstrated in simultaneous PET/MRI for brain imaging, including prospective MC (motion correction) of MRI. For both modalities, a reduction in motion-induced artifacts has been obtained after MC. The system proves the robust motion tracking. During the clinical study of 94 pediatric patients, 88 scans had no rejected tracking estimates, while five scans had just minor dropouts. The motion tracking capabilities, from small respiratory motions to the largest motion, possible in the MRI head coil have been proven.

Marker placement devices [13] could eliminate the greatest absolute difference by 8° for the hip flexion angle. The deficiencies that these devices bring into MoCap are: Primarily, the patient should stand still for several minutes, a difficult task for clinical settings, and second, they cannot be tested on patients whose orthopedic structures have been modified, either by surgery or children with cerebral palsy, that grew up.

6. Conclusions and perspectives

In most cases, motion capture can be exponentially aided by altering the environment, rather than altering the devices or the software involved in the process. For instance, the lighting, for cameras, the electromagnetic fields, for sensors, the cleanliness of the air, or water, for particle reduction, homogenizing heat, in favor of the infrared sensors, and so on.

The use of machine learning can attract breakthroughs in pose estimation due to the fact that having an already existing database with all the motion constraints, scales down the error space, and broadens not only the result fidelity, but also the decision-making processes.

Depending on the flexibility of the MoCap data, it can be used for analysis, processing, or it can be sent to other destinations, and be used as control data, for example in robotics and automation control, hologram piloting, or even anatomic motion rehabilitation through electric muscle stimulation.

The number of problems, regarding MoCap could never be determined as of yet, due to the novelty of the domain itself, which is why, the approach domains and equipment is growing wider even as we speak, and there will be no exact number of known conventional formulas to work with.

By the number and domical expanse to which MoCap processes have been brought, by now, the best MoCap approach, to each destination can only be determined by the technical skills of the scientists, and their power of choice.

References

1. Nicolella DP, Torres-Ronda L, Saylor KJ, Schelling X (2018)
2. Menolotto M, Komaris D-S, Tedesco S, O'Flynn B, Walsh M. *M Sensors*. 2020;
3. Fuller, R. Buckminster. *Critical Path*. New York: St. Martin's Press, 1981 (1981).
4. Naeemabadi M, Dinesen B, Andersen OK, Hansen J (2018)
5. Sila-Nowicka K, Thakuriah P (2019)
6. Sun C, Sun P, Wang P (2015) - (2018)
7. Dahan A, Noy L, Hart Y, Mayo A, Alon U (2016)
8. Westlund JK, D'Mello SK, Olney AM (2015)
9. MALTZ, M. *Psycho-cybernetics: a new way to get more living out of life*. N. Hollywood, Calif, Wilshire Book. (1976)
10. Tits M, Tilmanne J, Dutoit T (2018)
11. Bernardina GRD, Cerveri P, Barros RML, Marins JCB, Silvatti AP (2016)
12. Slipsager JM, Ellegaard AH, Glimberg SL, Paulsen RR, Tisdall MD, Wighton P, et al. (2019)
13. Gasparutto X, Wegrzyk J, Rose-Dulcina K, Hannouche D, Armand S (2020)