UGA Université Grenoble Alpes

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ GRENOBLE ALPES

Spécialité : **Informatique**

Arrêtée ministériel : 25 mai 2016

Présentée par

David PAGNON (doctorant)

Thèse dirigée par **Lionel REVERET (directeur)**
et codirigéee par **Mathieu DOMALAIN (co-encadrant)**

préparée au sein du **Laboratoire Jean Kuntzmann**
dans **l'École Doctorale l'École Doctorale Mathématiques, Sciences et technologies de l'Information, Informatique**

"Design, evaluation, and application of a workflow for biomechanically consistent markerless kinematics in sports"

"Conception, évaluation, et application d'une méthode biomécaniquement cohérente de cinématique sans marqueurs en sport"

Thèse soutenue publiquement le **"Date de soutenance"**,
devant le jury composé de :

Président
Laboratoire, Président

Rapporteur
Laboratoire, Rapporteur

Examinateur
Laboratoire, Examinateur

Lionel REVERET
INRIA Grenoble, Directeur de thèse

Mathieu DOMALAIN
Institut Pprime, Co-Encadrant de thèse

Invité
Laboratoire, Invité

SIGILLUM UNIVERSITATIS GRATIANOPOLITANE 1339

"Dédicace"

Remerciements

Remerciements possibles ici.

Abstract

Abstract.

Résumé

Résumé.

Contents

Introduction générale

Introduction générale.

1
State of the art

This chapter deals with the available prospects in motion capture (MoCap) in sports. It first present the standard marker-based (opto-electronic) systems for motion analysis and their limits. It briefly introduces some alternatives offered by Inertial Measurement Units (IMUs) or dept-field cameras. It then details the advent of markerless camera systems, which has been possible thanks to progress in machine learning. From 2D pose estimation to 3D joint angle determination, this is a new field which opens up new possiblities for motion analysis in a sports context.

This chapter is an up-to-date and slightly more detailed version of the introduction of the previously published paper "Pose2Sim: An End-to-End Workflow for 3D Markerless Sports Kinematics—Part 1: Robustness" [Pagnon2021]

Sommaire

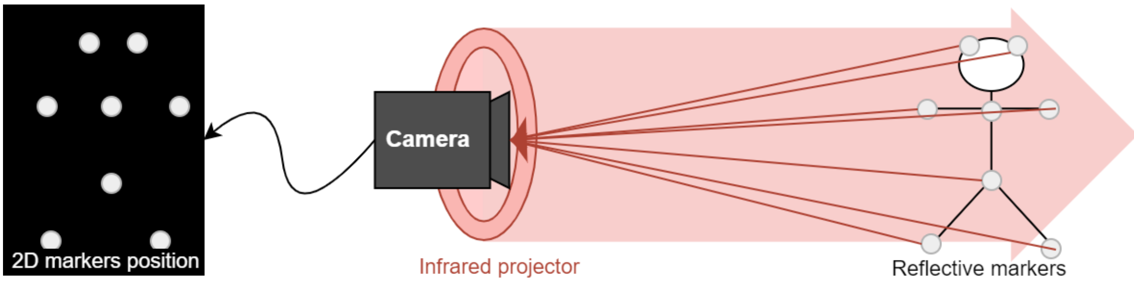
1.1 Overall context of kinematics in sports

As coaching athletes implies observing and understanding their movements, motion capture (MoCap) is essential in sports. It helps improving movement efficiency, preventing injuries, or predicting performances. For the last few decades, marker-based systems have been considered the best choice for the analysis of human movement, when regarding the trade-off between ease of use and accuracy. However, these methods have proven to be much more challenging in a sports context than in a laboratory setting, and to be generally inappropriate [Mündermann2006].

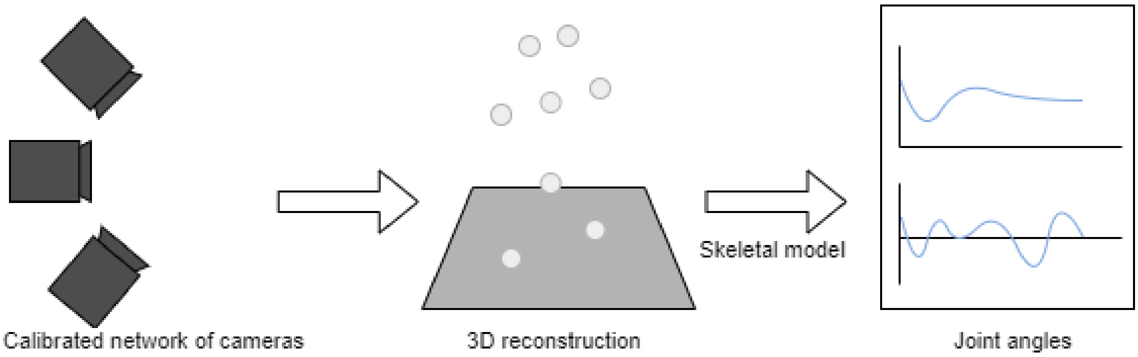
1.1.1 Marker-based systems

Marker-based systems use a network of opto-electronic cameras. Each of these cameras are surrounded by a crown of infrared LEDs, which projects light toward the subject. The subject wears special markers, reflecting light back to the camera. The camera usually pre-processes the image to make it binary, and only outputs the coordinates of the detected marker candidates (Figures 1.1a).

If calibrated, using a network of these cameras allows for triangulating the 2D coordinates. Calibration involves knowing the cameras' intrinsic properties (such as focal length, optical center, distortion) as well as their extrinsic properties (their positions as regards to the global coordinate system.) See Chapter 2.2 on 3D reconstruction for more details. The reconstructed 3D marker positions are then used to optimize the posture of a physically consistent skeleton, scaled to each individual subject. In particular, this allows for obtaining 3D joint angles at each point in time.



(a) An opto-electronic camera is traditionnaly surrounded by a crown of infrared LEDs, projecting light toward the subject. The subject wears markers, which reflect light back to the camera.



(b) Once calibrated, a network of these cameras allows for 3D reconstruction of marker positions. Marker coordinates are then used to infer the posture of the subject.

Figure 1.1: Principles of marker-based motion capture. (Figures 1.1a) presents the functionning of an opto-electronic camera. (Figures 1.1b) shows how a network of calibrated motion capture cameras helps obtaining joint angles.

Yet, reflective marker-based camera systems are complex to set up, are time-consuming, and are very expensive. They also require specific lightning conditions, and involve cumbersome cabling. Moreover, markers may fall off the body of the participant due to sharp accelerations or sweat. They can hinder the natural movement of athletes, which is likely to affect their warm-up, focus, and safety. While the accuracy of landmark location is claimed to be sub-millimetric in marker-based methods [Topley2020], marker placement is tedious, intrusive, prone to positioning variability from the operator [Tsushima2003], and subject to skin movement artifacts, especially on soft tissues. Della Croce et al. found out that inter-operator variations in marker placements range from 13 to 25 mm, which can propagate up to 10° in joint angle prediction [Gorton2009, della Croce1999]. For example, tissue artifacts account for up to a 2.5 cm marker displacement at the thigh, which can cause as much as a 3° error in knee joint angles tissues [Benoit2015, Cappozzo1995]. Joint positions must be calculated explicitly in marker-based methods, which introduces more variability: these errors range up to 5 cm, which can contribute up to 3° of error in lower limb joint angles [Leboeuf2019]. Nevertheless, since marker-based methods benefit from decades of research, they are still considered as the reference method for motion capture.

1.1.2 IMU and RGBD systems

Consequently, other approaches based on alternative technologies have been investigated over the past years.

For instance, wearable Inertial Measurement Units (IMUs) can be placed on an athlete's limbs. IMUs are made of an accelerometer, a gyroscope, and a magnetomete. The accelerometer measures the linear acceleration, the gyroscope measures the rotational speed, and the magnetometer measures the orientation of the earth magnetic field. Fusing and integrating these signals allows for the determination of their 3D orientations. The orientation of the athlete's limbs can then be used in combination with a skeletal model to infer the posture of the subject (Figures 1.2).

IMUs offer the advantages of getting away from all camera-related issues. They are inexpensive, they do not involve any complex setup and calibration, the field of view is larger, they are not sensitive to self- and gear-occlusions, they can be operated outside of a controlled environment, and they can work in real-time [Johnston2019, Chambers2015]. They still have the drawback of being an external equipment to wear, of involving high technical skills, and of being sensitive to ferromagnetic disturbances. Above all, they are exposed to drift over time and need to be calibrated every few minutes. Joint angle accuracy is relatively good in the flexion/extension plane, but less so in other rotational planes where errors are greater than 5° for most motions [Zhang2013, Rekant2022]. Moreover, they are not suitable for joint positions assessment, since these are obtained through multiple integrations of the original signal [Ahmad2013].

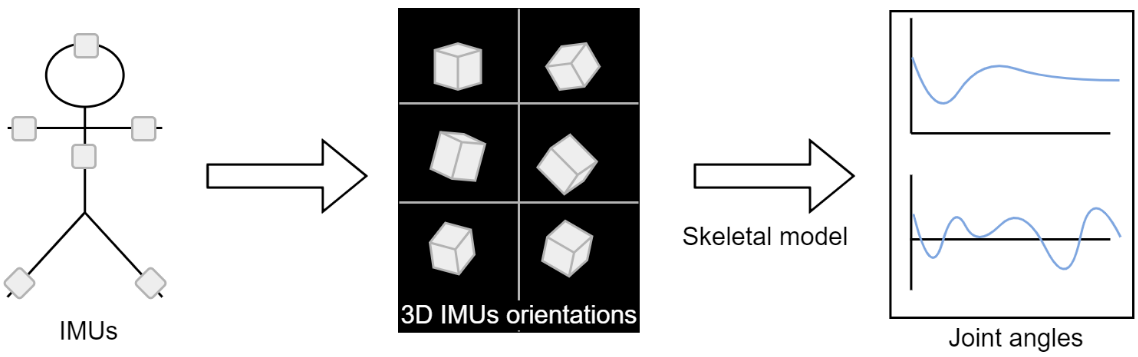


Figure 1.2: The search for "deep learning 3D human pose estimation" (dots) fits an exponential curve (line). The search produced less than 100 results until 2015. Over the course of 5 years, the number has reached almost 750 of them, and is now well over a 1000 per year.

Another approach involves depth-field cameras (RGBD), which give access to a 2.5D world (only the depth of the front facing plane of view is measured), or even to full 3D with a network of few RGBD cameras [Carraro2017, Choppin2013, Colombel2020]. On the other hand, these cameras suffer from a sensitivity to lightning conditions, they work at lower framerates, and they are short-range [Han2013].

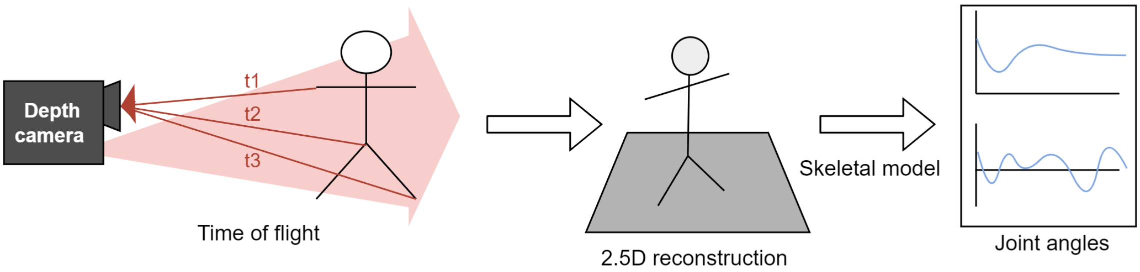


Figure 1.3: The search for "deep learning 3D human pose estimation" (dots) fits an exponential curve (line). The search produced less than 100 results until 2015. Over the course of 5 years, the number has reached almost 750 of them, and is now well over a 1000 per year.

1.1.3 Markerless systems

A recent breakthrough has come from Computer Vision. The explosion of deep-learning based methods from 2D camera videos, for which the research has skyrocketed around 2016 [Wang2021], is related to the increase in storage capacities and huge improvements in GPU computing. A search on ScienceDirect for "deep learning 3D human pose estimation" produced fewer than 100 papers per year until 2015, and the number is now reaching almost 750 over the span of 5 years, fitting an exponential curve (Figures 1.4).

It has rekindled interest from the Biomechanics community towards image-based motion analysis, which is where it all started with the invention of chronophotography in the 19th century by Marey in France, and Muybridge in the USA [Baker2007]. Currently, two approaches coexist in human and animal motion analysis: on the one hand, computer vision using deep-learning techniques mostly focus on joint positions only; while the interest of biomechanics lies in kinematics, that involves joint angles. One of the main current challenges is to bridge the gap between these two worlds, and to take advantage of deep-learning technologies for kinematic analysis [Cronin2021, Seethapathi2019].

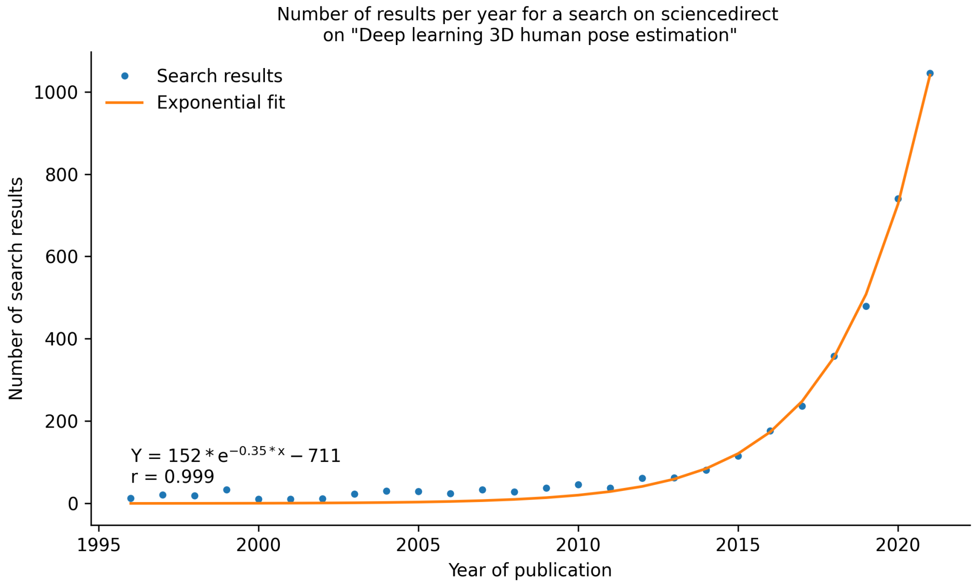


Figure 1.4: The search for "deep learning 3D human pose estimation" (dots) fits an exponential curve (line). The search produced less than 100 results until 2015. Over the course of 5 years, the number has reached almost 750 of them, and is now well over a 1000 per year.

1.2 2 dimensional analysis

1.2.1 2D pose estimation

The most well-known off-the-shelf 2D human pose estimation solutions are OpenPose [Cao2019], and to a lesser extent AlphaPose [Fang2017]. While both show similar results, OpenPose has the advantage of being a bottom-up approach, whose computation time does not increase with the number of persons detected [Cao2019]. A bottom-up approach first detects all available joint keypoints, and then associates them to the right persons; while a top-bottom approach first detects bounding boxes around each person, and then finds joint keypoints inside of them. OpenPose has been trained on the CMU Panoptic Dataset [Joo2015], with 511 synchronized videos of multiple people in motion, alone or engaged in social activities. Another 2D pose estimation toolbox is DeepLabCut [Mathis2018], which was initially intended for markerless animal pose estimation, but which can be custom trained for the detection of any human or not human keypoint with a relatively small dataset. All of these tools are open-source. Other approaches have shown even better results on evaluation datasets (see review [Chen2020]), but they are generally slower and not as widespread.

1.2.2 2D kinematics from 2D pose estimation

Some authors bridge 2D pose estimation to more biomechanically inspired goals, such as gait kinematics analysis. Kidzinski et al. present a toolbox for quantifying gait pathology that runs in a Google Colab [Kidziński2020]. Stenum et al. evaluate gait kinematics calculated from OpenPose input concurrently with a marker-based method. Mean absolute error of hip, knee and ankle sagittal angles were 4.0°, 5.6° and 7.4° [Stenum2021]. Liao et al. have not released their code, but they use OpenPose outputs to train a model invariant to view [Liao2020]. Viswakumar et al. perform direct calculation of the knee angle from an average phone camera processed by OpenPose [Viswakumar2019]. They show that OpenPose is robust to challenging clothing such as large Indian pants, as well as to extreme lightning conditions. Other sports activities have been investigated, such as lower body kinematics of vertical jump [Drazan2021] or underwater running [Cronin2019]. Both works train their own model with DeepLabCut. Serrancoli et al. fuse OpenPose and force sensors to retrieve joint dynamics in a pedaling task [Serrancolí2020].

1.3 3 dimensional analysis

1.3.1 3D pose estimation

There are a lot of different approaches for markerless 3D human pose estimation and listing them all is beyond our scope (see review [Wang2021]]). Some more ancient ones are not based on deep-learning and require specific lightning and background conditions, such as visual-hull reconstruction [Ceseracciu2014]. Some directly lift 3D from a single 2D camera (see review [Chen2020]), with different purposes: one estimates the positions of a set of keypoints around the joint instead of determining only the joint center keypoint, so that axial rotation along the limb is solved [Fisch2020]; SMPL and its sequels retrieve not only joint positions and orientations, but also body shape parameters [Loper2015]; while XNect primarily focuses on real time [Mehta2020]. A few approaches even strive to estimate 3D dynamics and contact forces from a 2D video input [Rempe2021, Li2019]. Rempe et al. solve occlusions from a 2D input [Rempe2020], but this remains a probabilistic guess that may be unsuccessful in case of unconventional positions of hidden limbs, whereas using more cameras would have given more trustworthy results . Haralabidis et al. fuse OpenPose results from a single monocular video and two IMU outputs, and solve kinematics of the upper body in OpenSim (an open-source biomechanical 3D analysis software [Delp2007, Seth2018]) in order to examine the effects of fatigue on boxing [Haralabidis2020].

Some research attempts to solve 3D pose estimation from a network of uncalibrated cameras, i.e., cameras whose extrinsic parameters (translation and rotation with respect to the coordinate system), intrinsic parameters (focal length, pixel size, etc.), and distortion coefficients are not known. It either uses 2D pose estimations of each view as visual cues to calibrate on [Takahashi2018], or an adversarial network that predicts views of other cameras, compares them to real views, and adjusts its calibration accordingly [Ershadi-Nasab2021]. Dong et al. recover 3D human motion from unsynchronized and uncalibrated videos of a repeatable movement found on internet videos (such as a tennis serve performed by a celebrity) [Dong2020]. Using uncalibrated videos is still a very experimental trend, that would require more research before being used in biomechanics.

We choose to focus on the methods that estimate 3D pose by triangulating 2D pose estimations from a network of multiple calibrated cameras. The classical evaluation metric is the MPJPE (Mean Per Joint Position Error), which is the average Euclidian distance between the estimated and the ground truth joint coordinate. Most methods take OpenPose as an input for triangulation, and more specifically the body_25 model. Labuguen et al. evaluate 3D joint positions of a pop dancer with a simple Direct Linear Transform triangulation (DLT [Hartley1997, Miller1980]) from 4 cameras [Labuguen2020]. Apart from the upper body for which error goes up to almost 700 mm, the average joint position error is about 100 mm. Nakano et al. examine three motor tasks (walking, countermovement jumping, and ball throwing), captured with 5 cameras and triangulated with the same methods, with a subsequent Butterworth filter [Nakano2019]. 47% of the errors are under 20 mm, 80% under 30 mm, and 10% are above 40 mm. The largest errors are mostly caused by OpenPose wrongly tracking a joint, for example by swapping the left and the right limb, what causes large errors up to 700 mm. This may be fixed either by using a better 2D pose estimator, or by using more cameras to reduce the impact of an error on a camera, or else by considering the temporal continuity in movement.

Slembrouck at al. go a step further and tackle the issue of limb swapping and of multiple persons detection [Slembrouck2020]. In case of multiple persons detection, one needs to make sure they triangulate the person detected on one camera to the same person detected on the other ones. Slembrouck et al. manage to associate persons across cameras by examining all the available triangulations for the neck and mid-hip joints: the persons are the same when the distance between the triangulated point and the line defined by the detected point and the camera center is below a certain threshold. They only focus on lower limb. Their first trial features a person running while being filmed by seven cameras, whereas their second one involves a person doing stationary movements such as squats while filmed by 3 cameras. After filtering, the average positional error in the first case is about 40 mm, and it is roughly 30 mm in the second case (less than 20 mm for the ankle joint). Other authors deal with the multiperson issue in a slightly different way [Bridgeman2019,Chu2021,Dong2019]. In average, if the detected persons are correctly associated and the limbs don't swap, the average joint position error for an OpenPose triangulation is mostly below 40 mm.

Some triangulation methods not based on OpenPose reach even better results on benchmarks, although it comes at the cost of either requiring heavy computations, or of being out of reach for non-expert in computer vision. The classic approach is to detect probability maps for each joint, to assume that the maximum probability is the actual 2D joint position, and then to triangulating these scalar positions. Instead of this, the main two state-of-the art methods directly perform a volumetric triangulation of the heatmaps, and only then take the maximum probability. By working this way, they keep all the information available for as possible for as long as possible. They manage to lower their MPJPE to about 20 mm [He2020, Iskakov2019].

1.3.2 3D kinematics from 3D pose estimation

Instead of just working on 3D joint positions, the issue of 3D markerless kinematics (i.e., gait parameters and joint angles) is starting to be tackled. Zago et al. evaluate gait parameters computed by triangulating 2 videos processed by OpenPose, and notice that straight gait direction, longer distance from subject to camera, and higher resolution make a big difference in accuracy [Zago2020]. D'Antonio at al. perform a simple triangulation of the OpenPose output of two cameras, and compute direct Euler angle calculations for the lower limb [D'Antonio2021]. They compare their results to IMU and point out that errors are higher for running than for walking, and are also rather inconsistent: up to 14°, although they can get down to 2 to 7° if the two cameras are set laterally rather than in the back of the subject. Theia3D, a recent commercial (and not open) solution, estimates the positions of a set of keypoints around the joint, and then uses a multi-body optimization approach to solve inverse kinematics [Kanko2021a, Kanko2021b]. They notice an offset in hip and ankle angles between their markerless system and the reference marker-based one, likely doe to different skeletal models. Once this offset is removed the root mean square error (RMSE) in lower limb roughly ranges between 2 and 8° for flexion/extension and abduction/adduction angles, and up to 11.6° for internal/external rotation. AniPose broadens the perspective to the kinematics of any human or animal with a DeepLabCut input, instead of OpenPose. They offer custom temporal filters, as well as spatial constraints on limb lengths [Karashchuk2021].

1.4 Statement of need

According to Atha [Atha1984], an ideal motion analysis system involves the collection of accurate information, the elimination of interference with natural movement, and the minimization of capture and analysis times. Yet, despite a marker-based system gives relatively accurate results, it requires placing markers on the body, which can hinder natural movement, and which are not it is hard to set up outdoors or in context, and it is strenuous to analyse.

Coaches usually investigate performance indicators and compare athletes, and compare with subjective visual observation However, despite these benefits, investing in technology has its pitfalls: the information gathered can be unhelpful, or inaccurate, or not easily interpretable, or not implementable in the context of sports [Windt2020].

The emergence of markerless kinematics opens up new possibilities. Indeed, the interest in deep-learning pose estimation neural networks has been growing fast since 2015 [Zheng2022], which makes it now possible to collect accurate and reliable kinematic data without the use of physical markers. OpenPose, for example, is a widespread open-source software which provides 2D joint coordinate estimations from videos. These coordinates can then be triangulated in order to produce 3D positions. Yet, when it comes to the biomechanical analysis of human motion, it is often more useful to obtain joint angles than their XYZ positions in space. Joint angles allow for better comparison among trials and individuals, and they represent the first step for other analysis such as inverse dynamics.

OpenSim is another widespread open-source software which helps compute 3D joint angles, usually from marker coordinates. It lets scientists define a detailed musculoskeletal model, scale it to individual subjects, and perform inverse kinematics with customizable biomechanical constraints. It provides other features such as net calculation of joint moments or individual muscle forces resolution, although this is out of the scope of our contribution.

The goal of Pose2Sim is to build a bridge between the communities of computer vision and biomechanics, by providing a simple and open-source pipeline connecting the two aforementioned state-of-the-art tools: OpenPose and OpenSim. Pose2Sim has already been used and tested in a number of situations (walking, running, cycling, balancing, swimming, boxing), and published in peer-reviewed scientific publications [Pagnon2021, Pagnon2022] assessing its robustness and accuracy. The combination of its ease of use, customizable characteristics, and high robustness and accuracy makes it promising, especially for "in-the-wild" sports movement analysis.

So far, little work has been done towards obtaining 3D angles from multiple views [Zheng2022]. However, two software applications are worth mentioning. Anipose [Karashchuk2021] proposes a Python open-source framework which allows for joint angle estimation with spatio-temporal constraints, but it is primarily designed for animal motion analysis. Theia3D [Kanko2021a] is a software application for human gait kinematics from videos. Although the GUI is more user friendly, it is not open-source nor easily customizable. Our results on inverse kinematics were deemed good when compared to marker-based ones. See [Pagnon2022] for more details on concurrent accuracy with other systems.

—

Currently, reference methods in sports analysis remain marker-based. These methods, also known as MoCap (motion capture) procedures, are mostly concerned with accuracy, despite the fact that marker placement hinders natural movement and is time consuming. Therefore, several markerless technologies are being examined to solve these issues. The main candidates are either based on Inertial Measurement Units (IMUs) [2,3], depth cameras [4,5,6], or a network of RGB cameras [7,8,9]. IMUs avoid all camera-related issues such as complex setup and calibration, potential self- and gear obstructions, and can operate in real time; however, they need to be worn by the athlete and are sensitive to drift over time, and to ferromagnetic disturbances. Depth cameras offer more information than RGB cameras but they hardly work in direct sunlight nor at a distance over 5 m [10]. On the other hand, a network of RGB cameras does not assume any particular environment, and it does not hinder the athlete's movement and focus, but it requires delicate calibration, complex setup, large storage space, and high computational capacities. The technology, however, is still maturing and some light-weight systems such as BlazePose [11] or UULPN [12] are being proposed, which can operate in real time on a mobile phone; however, they are still not quite as accurate as required for quantitative motion analysis.

We focus on the latter approach, and more specifically on methods triangulating 2D joint center estimations from a network of several calibrated RGB cameras. The most common evaluation metric is the Mean Per Joint Position Error (MPJPE), which is the average Euclidian distance between the estimated joint coordinate and its ground truth. A large part of studies investigating 3D joint center estimation choose to triangulate the output of OpenPose [13], a deep-learning algorithm estimating 2D joint coordinates from videos. Their MPJPE usually lies between 30 and 40 mm [14,15,16]. Ankle MPJPEs are within the margin of error of marker-based technologies (1–15 mm), whereas knee and hip MPJPEs are greater (30–50 mm). These errors are systematic and likely due to "ground-truth" images being mislabeled in the training dataset [17]. Triangulation from other 2D deep-learning algorithms (such as AlphaPose [18] and DeepLabCut [19]) have also been compared [17]. AlphaPose results are similar to OpenPose's; however, DeepLabCut errors are substantially higher.

Numerous studies have focused on the accuracy of 3D joint center estimation, but far fewer have examined 3D joint angle estimation. D'Antonio et al. computed direct flexion-extension angles for the lower limb from two cameras processed with OpenPose [20]. Range of Motion (ROM) errors lay between 2.8° and 14.1°. Wade et al. calculated frontal and sagittal knee and hip angles with OpenPose, AlphaPose, and DeepLabCut [21]. They deemed the method accurate enough for assessing step length and velocity, but not for joint angle analysis. AniPose offers a toolkit for triangulating 2D poses from DeepLabCut [22]. To our knowledge, it has only been concurrently validated for index finger angles in the sagittal plane, resulting in a root-mean-square error of 7.5 degrees [23]. Theia, a commercially available software package for markerless analysis, uses its own patent-protected 2D pose estimator and triangulation procedure, and runs a skeletal model to constrain the results to physically consistent poses and movements [24]. Their root-mean-square error (RMSE) compared to a marker-based method ranged between 2.6° and 13.2°.

1.5 Exemples

1.5.1 Tableaux

Générateur en ligne ici.

Un exemple de tableau générée par cet outil est présenté Table 1.1.

<table>
<tr><th></th><th>A</th><th>B</th><th>C</th></tr>
<tr><td>α</td><td colspan="3">fusion</td></tr>
<tr><td>β</td><td rowspan="2"></td><td>1</td><td>2</td></tr>
<tr><td>Δ</td><td>3</td><td>4</td></tr>
</table>

Tableau 1.1: Exemple de tableau

2
Theoretical framework

Résumé du chapitre possible ici.

Sommaire

2.1 Machine learning

2.1.1 Principles and history

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

2.1.2 Application to object detection and localization

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

2.1.3 Pose detection: Different architectures, different models, different datasets

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

2.2 3D reconstruction

2.2.1 Pinhole camera model

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

2.2.2 Calibration

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be

written in of the original language. There is no need for special content, but the length of words should match the language.

2.2.3 Triangulation

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

3

Proposed solution: Pose2Sim python package

Résumé du chapitre possible ici.

Sommaire

3.1 Introduction to the workflow

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

3.2 2D pose detection

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

3.3 Pose2Sim core

3.3.1 Tracking of the person viewed by the most cameras

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

3.3.2 Triangulating by weighted direct linear transform

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

3.3.3 Filtering

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be

written in of the original language. There is no need for special content, but the length of words should match the language.

3.4 Pose2Sim skeletal model

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

3.5 Limitations and perspectives

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

3.6 Helper functions and vizualisation tools

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

4

Pose2Sim robustness

Résumé du chapitre possible ici.

Sommaire

4.1 Introduction

4.1.1 Robustness definition

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

4.1.2 Assessing robustness

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

4.2 Methods

4.2.1 Experimental setup

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

4.2.2 Participant and protocol

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

4.2.3 Challenging robustness

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be

written in of the original language. There is no need for special content, but the length of words should match the language.

4.2.4 Statistical analysis

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

4.3 Results

4.3.1 Data collection and 2D pose estimation

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

4.3.2 Pose2Sim tracking, triangulation, and filtering

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

4.3.3 Relevance, repeatability and robustness of angles Results

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

4.4 Discussion

4.4.1 Pose2Sim

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

4.4.2 Relevance, repeatability and robustness

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

4.4.3 Limits and perspectives

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

5
Pose2Sim accuracy

Résumé du chapitre possible ici.

Sommaire

5.1 Introduction

5.1.1 State of the art

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

5.1.2 Assessing accuracy

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

5.2 Methods

5.2.1 Data collection

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

5.2.2 Markerless analysis

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

5.2.3 Marker-based analysis

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be

written in of the original language. There is no need for special content, but the length of words should match the language.

5.2.4 Statistical analysis

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

5.3 Results

5.3.1 Concurrent validation

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

5.3.2 Comparison with other systems

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

5.4 Discussion

5.4.1 Strengths of Pose2Sim and of markerless kinematic

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

5.4.2 Limits and perspectives

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

5.5 Conclusions

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

6

Pose2Sim in suboptimal conditions: Application to boxing

Résumé du chapitre possible ici.

Sommaire

6.1 Objectives

6.1.1 Key Performance Indicators in boxing

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

6.1.2 Limits of research-grade systems in competitions

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

6.1.3 Objectives

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

6.2 Methods

6.2.1 4 conditions

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

6.2.2 Pose-calibration on ring dimensions

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be

written in of the original language. There is no need for special content, but the length of words should match the language.

6.2.3 Post-synchronization on 2D movement speeds

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

6.2.4 GoPro spatio-temporal base into Qualysis’

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

6.2.5 Statistical analysis

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

6.3 Results

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

6.4 Discussion

6.4.1 Equipment and protocol vs. pose estimation model

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

6.4.2 Pros and cons of different systems

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

7

OpenPose and DeepLabCut together with Pose2Sim: Application to BMX racing

Résumé du chapitre possible ici.

Sommaire

7.1 Introduction

7.1.1 The start in BMX racing

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

7.2 Methods

7.2.1 Material and protocol

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

7.2.2 Pilot inverse kinematics

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

7.2.3 Bike inverse kinematics

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

7.2.4 Joined pilot and bike inverse kinematics

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like "Huardest gefburn"? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be

written in of the original language. There is no need for special content, but the length of words should match the language.

7.3 Results

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

7.4 Discussion

7.4.1 On these data

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

7.4.2 Limits and perspectives

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

Conclusion générale

Conclusion ici.

Bibliography

[Ahmad2013] Norhafizan Ahmad, Raja Ariffin Raja Ghazilla, Nazirah M. Khairi et Vijayabaskar Kasi. *Reviews on Various Inertial Measurement Unit (IMU) Sensor Applications*. International Journal of Signal Processing Systems, pages 256–262, 2013.

[Atha1984] J Atha. *Current techniques for measuring motion*. Applied ergonomics, vol. 15, no. 4, pages 245–257, 1984.

[Baker2007] Richard Baker. *The history of gait analysis before the advent of modern computers*. Gait and Posture, vol. 26, no. 3, pages 331–342, 9 2007.

[Benoit2015] D. L. Benoit, M. Damsgaard et M. S. Andersen. *Surface marker cluster translation, rotation, scaling and deformation: Their contribution to soft tissue artefact and impact on knee joint kinematics*. Journal of Biomechanics, vol. 48, no. 10, pages 2124–2129, 7 2015.

[Bridgeman2019] Lewis Bridgeman, Marco Volino, Jean-Yves Guillemaut et Adrian Hilton. *Multi-Person 3D Pose Estimation and Tracking in Sports*. pages 2487–2496, Long Beach, CA, USA, 6 2019. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), IEEE.

[Cao2019] Zhe Cao, Gines Hidalgo, Tomas Simon, Shih-En Wei et Yaser Sheikh. *OpenPose: realtime multi-person 2D pose estimation using Part Affinity Fields*. IEEE transactions on pattern analysis and machine intelligence, vol. 43, no. 1, pages 172–186, 2019.

[Cappozzo1995] A Cappozzo, F Catani, U Della Croce et A Leardini. *Position and orientation in space of bones during movement: anatomical frame definition and determination*. Clinical Biomechanics, vol. 10, no. 4, pages 171–178, 6 1995.

[Carraro2017] Marco Carraro, Matteo Munaro, Jeff Burke et Emanuele Menegatti. *Real-time marker-less multi-person 3D pose estimation in RGB-Depth camera networks*. arXiv:1710.06235 [cs], 10 2017. arXiv: 1710.06235.

[Ceseracciu2014] Elena Ceseracciu, Zimi Sawacha et Claudio Cobelli. *Comparison of Markerless and Marker-Based Motion Capture Technologies through Simultaneous Data Collection during Gait: Proof of Concept*. PLoS ONE, vol. 9, no. 3, page e87640, 3 2014.

[Chambers2015] Ryan Chambers, Tim J Gabbett, Michael H Cole et Adam Beard. *The use of wearable microsensors to quantify sport-specific movements*. Sports medicine, vol. 45, no. 7, pages 1065–1081, 2015.

[Chen2020] Yucheng Chen, Yingli Tian et Mingyi He. *Monocular human pose estimation: A survey of deep learning-based methods*. Computer Vision and Image Understanding, vol. 192, page 102897, 3 2020.

[Choppin2013] Simon Choppin et Jonathan Wheat. *The potential of the Microsoft Kinect in sports analysis and biomechanics*. Sports Technology, vol. 6, no. 2, pages 78–85, 5 2013.

[Chu2021] Hau Chu, Jia-Hong Lee, Yao-Chih Lee, Ching-Hsien Hsu, Jia-Da Li et Chu-Song Chen. *Part-Aware Measurement for Robust Multi-View Multi-Human 3D Pose Estimation and Tracking*. page 10, 2021.

[Colombel2020] Jessica Colombel, Vincent Bonnet, David Daney, Raphael Dumas, Antoine Seilles et François Charpillet. *Physically Consistent Whole-Body Kinematics Assessment Based on an RGB-D Sensor. Application to Simple Rehabilitation Exercises*. Sensors, vol. 20, no. 10, page 2848, 5 2020.

[Cronin2019] Neil J. Cronin, Timo Rantalainen, Juha P. Ahtiainen, Esa Hynynen et Ben Waller. *Markerless 2D kinematic analysis of underwater running: A deep learning approach*. Journal of Biomechanics, vol. 87, pages 75–82, 4 2019.

[Cronin2021] Neil J. Cronin. *Using deep neural networks for kinematic analysis: challenges and opportunities*. Journal of Biomechanics, page 110460, 5 2021.

[D'Antonio2021] Erika D'Antonio, Juri Taborri, Ilaria Mileti, Stefano Rossi et Fabrizio Patane. *Validation of a 3D Markerless System for Gait Analysis based on OpenPose and Two RGB Webcams*. IEEE Sensors Journal, pages 1–1, 2021.

[della Croce1999] U. della Croce, A. Cappozzo et D. C. Kerrigan. *Pelvis and lower limb anatomical landmark calibration precision and its propagation to bone geometry and joint angles*. Medical and Biological Engineering and Computing, vol. 37, no. 2, pages 155–161, 3 1999.

[Delp2007] Scott L Delp, Frank C Anderson, Allison S Arnold, Peter Loan, Ayman Habib, Chand T John, Eran Guendelman et Darryl G Thelen. *OpenSim: open-source software to create and analyze dynamic simulations of movement*. IEEE transactions on biomedical engineering, vol. 54, no. 11, pages 1940–1950, 2007.

[Dong2019] Junting Dong, Wen Jiang, Qixing Huang, Hujun Bao et Xiaowei Zhou. *Fast and Robust Multi-Person 3D Pose Estimation From Multiple Views*. pages 7784–7793, Long Beach, CA, USA, 6 2019. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE.

[Dong2020] Junting Dong, Qing Shuai, Yuanqing Zhang, Xian Liu, Xiaowei Zhou et Hujun Bao. *Motion Capture from Internet Videos*. In Andrea Vedaldi, Horst Bischof, Thomas Brox et Jan-Michael Frahm, editeurs, Computer Vision – ECCV 2020, volume 12347, pages 210–227. Springer International Publishing, Cham, 2020.

[Drazan2021] John F. Drazan, William T. Phillips, Nidhi Seethapathi, Todd J. Hullfish et Josh R. Baxter. *Moving outside the lab: Markerless motion capture accurately quantifies sagittal plane kinematics during the vertical jump*. Journal of Biomechanics, vol. 125, page 110547, 8 2021.

[Ershadi-Nasab2021] Sara Ershadi-Nasab, Shohreh Kasaei et Esmaeil Sanaei. *Uncalibrated multi-view multiple humans association and 3D pose estimation by adversarial learning*. Multimedia Tools and Applications, vol. 80, no. 2, pages 2461–2488, 1 2021.

[Fang2017] Hao-Shu Fang, Shuqin Xie, Yu-Wing Tai et Cewu Lu. *RMPE: Regional Multi-person Pose Estimation*. pages 2353–2362, Venice, 10 2017. 2017 IEEE International Conference on Computer Vision (ICCV), IEEE.

[Fisch2020] Martin Fisch et Ronald Clark. *Orientation Keypoints for 6D Human Pose Estimation*. arXiv:2009.04930 [cs], 9 2020. arXiv: 2009.04930.

[Gorton2009] George E. Gorton, David A. Hebert et Mary E. Gannotti. *Assessment of the kinematic variability among 12 motion analysis laboratories*. Gait and Posture, vol. 29, no. 3, pages 398–402, 4 2009.

[Han2013] Jungong Han, Ling Shao, Dong Xu et Jamie Shotton. *Enhanced Computer Vision With Microsoft Kinect Sensor: A Review*. IEEE Transactions on Cybernetics, vol. 43, no. 5, pages 1318–1334, 10 2013. event: IEEE Transactions on Cybernetics.

[Haralabidis2020] Nicos Haralabidis, David John Saxby, Claudio Pizzolato, Laurie Needham, Dario Cazzola et Clare Minahan. *Fusing Accelerometry with Videography to Monitor the Effect of Fatigue on Punching Performance in Elite Boxers*. Sensors (Basel, Switzerland), vol. 20, no. 20, 10 2020.

[Hartley1997] Richard I. Hartley et Peter Sturm. *Triangulation*. Computer Vision and Image Understanding, vol. 68, no. 2, pages 146–157, 11 1997.

[He2020] Yihui He, Rui Yan, Katerina Fragkiadaki et Shoou-I Yu. *Epipolar Transformers*. pages 7776–7785. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 6 2020. ISSN: 2575-7075.

[Iskakov2019] Karim Iskakov, Egor Burkov, Victor Lempitsky et Yury Malkov. *Learnable Triangulation of Human Pose*. pages 7717–7726, Seoul, Korea (South), 10 2019. 2019 IEEE/CVF International Conference on Computer Vision (ICCV), IEEE.

[Johnston2019] William Johnston, Martin O'Reilly, Rob Argent et Brian Caulfield. *Reliability, validity and utility of inertial sensor systems for postural control assessment in sport science and medicine applications: a systematic review*. Sports Medicine, vol. 49, no. 5, pages 783–818, 2019.

[Joo2015] Hanbyul Joo, Hao Liu, Lei Tan, Lin Gui, Bart Nabbe, Iain Matthews, Takeo Kanade, Shohei Nobuhara et Yaser Sheikh. *Panoptic Studio: A Massively Multiview System for Social Motion Capture*. pages 3334–3342. 2015 IEEE International Conference on Computer Vision (ICCV), 12 2015. ISSN: 2380-7504.

[Kanko2021a] Robert M. Kanko, Elise Laende, W. Scott Selbie et Kevin J. Deluzio. *Intersession repeatability of markerless motion capture gait kinematics*. Journal of Biomechanics, vol. 121, page 110422, 5 2021.

[Kanko2021b] Robert M. Kanko, Elise K. Laende, Elysia M. Davis, W. Scott Selbie et Kevin J. Deluzio. *Concurrent assessment of gait kinematics using marker-based and markerless motion capture*. Journal of Biomechanics, page 110665, 8 2021.

[Karashchuk2021] Pierre Karashchuk, Katie L Rupp, Evyn S Dickinson, Sarah Walling-Bell, Elischa Sanders, Eiman Azim, Bingni W Brunton et John C Tuthill. *Anipose: a toolkit for robust markerless 3D pose estimation*. Cell reports, vol. 36, no. 13, page 109730, 2021.

[Kidziński2020] Łukasz Kidziński, Bryan Yang, Jennifer L. Hicks, Apoorva Rajagopal, Scott L. Delp et Michael H. Schwartz. *Deep neural networks enable quantitative movement analysis using single-camera videos*. Nature Communications, vol. 11, no. 1, page 4054, 12 2020.

[Labuguen2020] Rollyn T. Labuguen, Wally Enrico M. Ingco, Salvador Blanco Negrete, Tonan Kogami et Tomohiro Shibata. *Performance Evaluation of Markerless 3D Skeleton Pose Estimates with Pop Dance Motion Sequence*. Rapport technique, 4 2020. DOI: 10.1101/2020.04.15.010702.

[Leboeuf2019] F. Leboeuf, J. Reay, R. Jones et M. Sangeux. *The effect on conventional gait model kinematics and kinetics of hip joint centre equations in adult healthy gait*. Journal of Biomechanics, vol. 87, pages 167–171, 4 2019.

[Li2019] Zongmian Li, Jiri Sedlar, Justin Carpentier, Ivan Laptev, Nicolas Mansard et Josef Sivic. *Estimating 3D Motion and Forces of Person-Object Interactions From Monocular Video*. pages 8632–8641, Long Beach, CA, USA, 6 2019. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE.

[Liao2020] Rijun Liao, Shiqi Yu, Weizhi An et Yongzhen Huang. *A model-based gait recognition method with body pose and human prior knowledge*. Pattern Recognition, vol. 98, page 107069, 2 2020.

[Loper2015] Matthew Loper, Naureen Mahmood, Javier Romero, Gerard Pons-Moll et Michael J. Black. *SMPL: a skinned multi-person linear model*. ACM Transactions on Graphics, vol. 34, no. 6, pages 1–16, 11 2015.

[Mathis2018] Alexander Mathis, Pranav Mamidanna, Kevin M. Cury, Taiga Abe, Venkatesh N. Murthy, Mackenzie Weygandt Mathis et Matthias Bethge. *DeepLabCut: markerless pose estimation of user-defined body parts with deep learning*. Nature Neuroscience, vol. 21, no. 9, pages 1281–1289, 9 2018.

[Mehta2020] Dushyant Mehta, Oleksandr Sotnychenko, Franziska Mueller, Weipeng Xu, Mohamed Elgharib, Pascal Fua, Hans-Peter Seidel, Helge Rhodin, Gerard Pons-Moll et Christian Theobalt. *XNect: real-time multi-person 3D motion capture with a single RGB camera*. ACM Transactions on Graphics, vol. 39, no. 4, page 82:82:1–82:82:17, 7 2020.

[Miller1980] Norman R. Miller, Robert Shapiro et Thomas M. McLaughlin. *A technique for obtaining spatial kinematic parameters of segments of biomechanical systems from cinematographic data*. Journal of Biomechanics, vol. 13, no. 7, pages 535–547, 1 1980.

[Mündermann2006] Lars Mündermann, Stefano Corazza et Thomas P. Andriacchi. *The evolution of methods for the capture of human movement leading to markerless motion capture for biomechanical applications*. Journal of NeuroEngineering and Rehabilitation, vol. 3, no. 1, page 6, 3 2006.

[Nakano2019] Nobuyasu Nakano, Tetsuro Sakura, Kazuhiro Ueda, Leon Omura, Arata Kimura, Yoichi Iino, Senshi Fukashiro et Shinsuke Yoshioka. *Evaluation of 3D markerless motion capture accuracy using OpenPose with multiple video cameras*. Rapport technique, 11 2019. DOI: 10.1101/842492.

[Pagnon2021] David Pagnon, Mathieu Domalain et Lionel Reveret. *Pose2Sim: An End-to-End Workflow for 3D Markerless Sports Kinematics—Part 1: Robustness*. Sensors, vol. 21, no. 19, 2021.

[Pagnon2022] David Pagnon, Mathieu Domalain et Lionel Reveret. Sensors, vol. 22, no. 7, 2022.

[Rekant2022] Julie Rekant, Scott Rothenberger et April Chambers. *Inertial measurement unit-based motion capture to replace camera-based systems for assessing gait in healthy young adults: Proceed with caution*. Measurement: Sensors, page 100396, 2022.

[Rempe2020] Davis Rempe, Leonidas J Guibas, Aaron Hertzmann, Bryan Russell, Ruben Villegas et Jimei Yang. *Contact and Human Dynamics from Monocular Video*. page 27, 2020.

[Rempe2021] Davis Rempe, Tolga Birdal, Aaron Hertzmann, Jimei Yang, Srinath Sridhar et Leonidas J Guibas. *HuMoR: 3D Human Motion Model for Robust Pose Estimation*. page 23, 2021.

[Seethapathi2019] Nidhi Seethapathi, Shaofei Wang, Rachit Saluja, Gunnar Blohm et Konrad P. Kording. *Movement science needs different pose tracking algorithms*. arXiv:1907.10226 [cs, q-bio], 7 2019. arXiv: 1907.10226.

[Serrancolí2020] Gil Serrancolí, Peter Bogatikov, Joana Palés Huix, Ainoa Forcada Barberà, Antonio J. Sánchez Egea, Jordi Torner Ribé, Samir Kanaan-Izquierdo et Antoni Susín. *Marker-Less Monitoring Protocol to Analyze Biomechanical Joint Metrics During Pedaling*. IEEE Access, vol. 8, pages 122782–122790, 2020. event: IEEE Access.

[Seth2018] Ajay Seth, Jennifer L. Hicks, Thomas K. Uchida, Ayman Habib, Christopher L. Dembia, James J. Dunne, Carmichael F. Ong, Matthew S. DeMers, Apoorva Rajagopal, Matthew Millard, Samuel R. Hamner, Edith M. Arnold, Jennifer R. Yong, Shrinidhi K. Lakshmikanth, Michael A. Sherman, Joy P. Ku et Scott L. Delp. *OpenSim: Simulating musculoskeletal dynamics and neuromuscular control to study human and animal movement*. PLOS Computational Biology, vol. 14, no. 7, page e1006223, 7 2018.

[Slembrouck2020] Maarten Slembrouck, Hiep Luong, Joeri Gerlo, Kurt Schütte, Dimitri Van Cauwelaert, Dirk De Clercq, Benedicte Vanwanseele, Peter Veelaert et Wilfried Philips. *Multiview 3D Markerless Human Pose Estimation from OpenPose Skeletons*. In Jacques Blanc-Talon, Patrice Delmas, Wilfried Philips, Dan Popescu et Paul Scheunders, editeurs, Advanced Concepts for Intelligent Vision Systems, volume 12002, pages 166–178. Springer International Publishing, Cham, 2020.

[Stenum2021] Jan Stenum, Cristina Rossi et Ryan T. Roemmich. *Two-dimensional video-based analysis of human gait using pose estimation*. PLoS Computational Biology, vol. 17, no. 4, 4 2021.

[Takahashi2018] Kosuke Takahashi, Dan Mikami, Mariko Isogawa et Hideaki Kimata. *Human Pose as Calibration Pattern: 3D Human Pose Estimation with Multiple Unsynchronized and Uncalibrated Cameras*. pages 1856–18567, Salt Lake City, UT, USA, 6 2018. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), IEEE.

[Topley2020] Matt Topley et James G. Richards. *A comparison of currently available optoelectronic motion capture systems*. Journal of Biomechanics, vol. 106, page 109820, 6 2020.

[Tsushima2003] Hitoshi Tsushima, Meg E Morris et Jennifer McGinley. *Test-Retest Reliability and Inter-Tester Reliability of Kinematic Data from a Three-Dimensional Gait Analysis System*. Journal of the Japanese Physical Therapy Association, vol. 6, no. 1, pages 9–17, 2003.

[Viswakumar2019] Aditya Viswakumar, Venkateswaran Rajagopalan, Tathagata Ray et Chandu Parimi. *Human Gait Analysis Using OpenPose*. pages 310–314. 2019 Fifth International Conference on Image Information Processing (ICIIP), 11 2019. ISSN: 2640-074X.

[Wang2021] Jinbao Wang, Shujie Tan, Xiantong Zhen, Shuo Xu, Feng Zheng, Zhenyu He et Ling Shao. *Deep 3D human pose estimation: A review*. Computer Vision and Image Understanding, page 103225, 5 2021.

[Windt2020] Johann Windt, Kerry MacDonald, David Taylor, Bruno D Zumbo, Ben C Sporer et David T Martin. *"To tech or not to tech?" A critical decision-making framework for implementing technology in sport*. Journal of Athletic Training, vol. 55, no. 9, pages 902–910, 2020.

[Zago2020] Matteo Zago, Matteo Luzzago, Tommaso Marangoni, Mariolino De Cecco, Marco Tarabini et Manuela Galli. *3D Tracking of Human Motion Using Visual Skeletonization and Stereoscopic Vision*. Frontiers in Bioengineering and Biotechnology, vol. 8, 2020.

[Zhang2013] Jun-Tian Zhang, Alison C Novak, Brenda Brouwer et Qingguo Li. *Concurrent validation of Xsens MVN measurement of lower limb joint angular kinematics*. Physiological measurement, vol. 34, no. 8, page N63, 2013.

[Zheng2022] Ce Zheng, Wenhan Wu, Taojiannan Yang, Sijie Zhu, Chen Chen, Ruixu Liu, Ju Shen, Nasser Kehtarnavaz et Mubarak Shah. *Deep learning-based human pose estimation: A survey*. arXiv, 2022.

List of Figures

List of Tables

A
Annexe 1 : Titre

Résumé de l'annexe ici.

A.1 Section 1

A.1.1 Sous section 1

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

A.1.2 Sous section 2

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

B
Annexe 2 : Titre

Résumé ici.

B.1 Section 1

B.1.1 Sous section 1

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

B.1.2 Sous section 2

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

C
Annexe 3 : Titre

Résumé ici.

C.1 Section 1

C.1.1 Sous section 1

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

C.1.2 Sous section 2

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

"Design, evaluation, and application of a workflow for biomechanically consistent markerless kinematics in sports"

"Conception, évaluation, et application d'une méthode biomécaniquement cohérente de cinématique sans marqueurs en sport"

Résumé

Ici ... résumé en français.

Mots-clés : Mots clés

Abstract

Ici ... résumé en anglais.

Keywords : markerless motion capture; sports performance analysis; kinematics; computer vision; openpose; opensim; python package