Marusic, A., Nguyen, S. M., and Tapus, A. (2023). Evaluating Kinect, OpenPose and BlazePose for Human Body Movement Analysis on a Low Back Pain Physical Rehabilitation Dataset. Companion of the 2023 ACM/IEEE International Conference on Human-Robot Interaction(587–591). Association for Computing Machinery.

Evaluating Kinect, OpenPose and BlazePose for Human Body Movement Analysis on a Low Back Pain Physical Rehabilitation Dataset

Aleksa Marusic
Autonomous Systems and Robotics Lab, Computer Science and System Engineering (U2IS), ENSTA Paris
Institut Polytechnique de Paris
aleksa.marusic@ensta-paris.fr

Sao Mai Nguyen
Autonomous Systems and Robotics Lab, Computer Science and System Engineering, ENSTA Paris, IP Paris
& Dep. Informatique, IMT Atlantique
nguyensmai@gmail.com

Adriana Tapus
Autonomous Systems and Robotics Lab, Computer Science and System Engineering (U2IS), ENSTA Paris
Institut Polytechnique de Paris
adriana.tapus@ensta-paris.fr

ABSTRACT

Analyzing human motion is an active research area, with various applications. In this work, we focus on human motion analysis in the context of physical rehabilitation using a robot coach system. Computer-aided assessment of physical rehabilitation entails evaluation of patient performance in completing prescribed rehabilitation exercises, based on processing movement data captured with a sensory system, such as RGB and RGB-D cameras. As 2D and 3D human pose estimation from RGB images had made impressive improvements, we aim to compare the assessment of physical rehabilitation exercises using movement data obtained from both RGB-D camera (Microsoft Kinect) and estimation from RGB videos (OpenPose and BlazePose algorithms). A Gaussian Mixture Model (GMM) is employed from position (and orientation) features, with performance metrics defined based on the log-likelihood values from GMM. The evaluation is performed on a medical database of clinical patients carrying out low back-pain rehabilitation exercises, previously coached by robot Poppy.

CCS CONCEPTS

• **Human-centered computing** → **Accessibility design and evaluation methods**; • **Computing methodologies** → *Mixture models*; *Classification and regression trees*; **Motion capture**; *Vision for robotics.*

KEYWORDS

Human Body Movement Analysis, Physical Rehabilitation, Human Skeleton Representation, Humanoid Robot, Motion Assessment, Robot Coach

ACM Reference Format:
Aleksa Marusic, Sao Mai Nguyen, and Adriana Tapus. 2023. Evaluating Kinect, OpenPose and BlazePose for Human Body Movement Analysis on a Low Back Pain Physical Rehabilitation Dataset. In . ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3568294.3580153

HRI '23 Companion, March 13–16, 2023, Stockholm, Sweden

ACM ISBN 978-1-4503-9970-8/23/03…$15.00
https://doi.org/10.1145/3568294.3580153

Figure 1: Setting of the system including a Microsoft Kinect v2 and an open source humanoid robot called Poppy

1 INTRODUCTION

Physical rehabilitation affects an increasing number of people and is usually prescribed to patients who suffer from certain disabilities, or need to restore functional abilities, usually after an injury or surgery. One of the most important conditions is low back pain (LBP), which is the leading cause of disability worldwide [3].

In rehabilitation programs, a clinician instructs patients on how to perform rehabilitation exercises and then monitors their performance in a clinical setting. Such treatment depends on the availability of physiotherapists. In order to increase flexibility, home-based rehabilitation is often used instead of clinic-based rehabilitation. In such systems, a physiotherapist makes a rehabilitation plan consisting of several recommended exercises. Patients then perform the exercises at home and visit the clinic periodically for progress assessment. However, the lack of supervision and timely feedback by a healthcare professional are considered the main factors for decreasing the engagement of the patients throughout the months-long repetition of physical exercises. Low motivation and poor supervision increase the chances of incorrect performance of the exercises by the patient, which slows down the recovery process and increases the risk of re-injury [14].

A solution for this problem can be a robot coach for physical rehabilitation exercises. The robot should be able to learn how to perform a rehabilitation exercise as well as to assess patients' movements. The system used in this paper is composed of an open-source humanoid robot called Poppy [10], and a depth camera (Microsoft Kinect v2), as detailed in [7, 8]. Figure 1 shows the setting of this system. The Kinect sensor is used to capture human motion of both the therapists and the patients. To be able to provide feedback, human motion analysis is therefore crucial.

Marusic, A., Nguyen, S. M., and Tapus, A. (2023). Evaluating Kinect, OpenPose and BlazePose for Human Body Movement Analysis on a Low Back Pain Physical Rehabilitation Dataset. Companion of the 2023 ACM/IEEE International Conference on Human-Robot Interaction(587–591). Association for Computing Machinery.

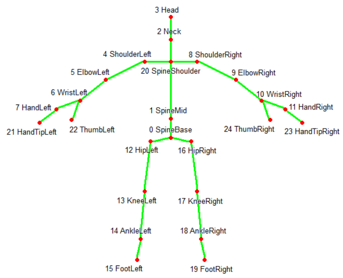


Figure 2: Kinect skeleton

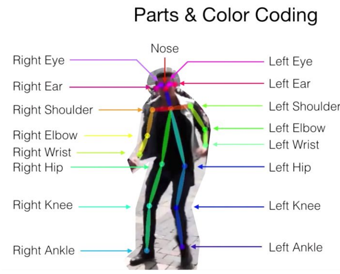


Figure 3: OpenPose skeleton

Analyzing human motion is a very active research topic today, with applications in several domains such as sports sciences [6], action and gesture recognition [2, 5, 9], and range-of-motion estimation [1]. Physiotherapists and GPs have extensive experience to classify a certain motion as correct or incorrect. Therefore, developing an automatic system for such a task is not easy due to a wide diversity of movements and a certain degree of subjectivity [1].

Obtaining precise movement data by motion sensors is crucial. In [14], the authors review different motion capture sensors used for rehabilitation exercise evaluation. One way is to use optical motion tracking systems, which place a set of markers on a patient's body, ankles, and key body parts, thus obtaining the precise position of each joint at each time. Although such systems are highly accurate and reliable, they are very expensive and need to be (re)installed every time the patient is doing any exercise. The other, non-invasive approach, is the usage of depth cameras, which provide both position and orientation of skeleton joints. These systems become very popular as they are cheap and easy to use. The most popular such system is Microsoft Kinect sensor, which has been used in most of the related works. Finally, a standard vision camera can be used as a motion sensor as well. Over the years, these cameras had a limit on the evaluation accuracy, but nowadays, with the huge development of computer vision and deep learning techniques, it is possible to estimate skeleton joints positions, and even orientations, from plain RGB images such as with algorithms OpenPose and BlazePose [13, 15].

Our work aims to evaluate the latter two sensor systems and investigates whether standard vision cameras, using OpenPose and BlazePose algorithms [13, 15], can have comparable results with motion captured by the Microsoft Kinect system.

The rest of the paper is structured as follows. Section 2 provides sensory data overview and the dataset used, while Section 3 describes the implemented method for rehabilitation assessment. The results are summarized in Section 4. The conclusions and discussions are part of Section 5.

2 DATASET

2.1 Skeleton data

A Human Pose Skeleton represents the orientation of a person in a graphical format. It is a set of coordinates that can be connected to describe the pose of the person. Points/coordinates in the skeleton are called joints or keypoints while the connections between them are pairs or limbs.

Figures 2 and 3 show joints of Kinect and OpenPose [1] skeletons.

2.2 Dataset

The Keraal dataset has been recorded within a long-term rehabilitation program, targeting low back pain. The data includes recordings from healthy subjects but, more importantly, of rehabilitation patients. The data is extracted from a 4 weeks evolution of each patient. The videos collected from patients were annotated by a physiotherapist, using the Anvil video annotation research tool. Each exercise was labeled as either correct or incorrect.

This section describes the protocol and rationale for how the Keraal dataset was created. It describes the participants included, the hardware, and the experimental protocol used in the dataset. It is made available on http://nguyensmai.free.fr/KeraalDataset.html[12].

2.2.1 Rehabilitation program. 31 patients, aged 18 to 70 years, were recruited in the double blind study. This prospective, centrally randomized, controlled, single-blind, and bi-centric study was conducted from October 2017 to May 2019. 12 patients suffering from low-back pain were included in the Robot Supervised Rehabilitation Group, and were asked by a humanoid robot coach to perform each of the three predefined exercises the best they can from its demonstration. The details on this clinical trial, including the patient care, the rehabilitation sessions, the robot coach, the inclusion and exclusion criteria, the characteristics of the patients, the efficiency of the care have been reported in Blanchard et al. [4].

2.2.2 Exercises and errors. A list of three exercises has been chosen in conjunction with therapists as common rehabilitation exercises that are also used for low-back pain treatment. Illustrations of these exercises can be seen as exercises 1, 2 and 3 of http://keraal.enstb.org/exercises.html. The 3 exercises are centered on spine stretching: a left rotation of the trunk followed by a right rotation, a left and right lateral bending of the trunk and a breathing exercise with the upper limbs flexed 90° at shoulder and elbow.

A list of common errors was defined in conjunction with the experience of the therapists from the CHRU Brest.

2.2.3 Participants. The dataset contains data from three groups :

- Group1A : Rehabilitation Patients : 14 recordings per exercise among the daily sessions of 6 patients were annotated.

[1] https://maelfabien.github.io/tutorials/open-pose/#run-openpose

- Group2A : Healthy participants: 6 healthy adults are free to execute each exercise correctly or with errors. 51 recordings per exercise were annotated.
- Group3 : Healthy participants : 3 healthy adults perform correctly the exercises and simulate the identified errors.

2.2.4 Sensor system. Using the Microsoft Kinect V2 sensor, we obtained the RGB video with the skeleton drawn, and the skeleton joint positions and orientations information. From the RGB videos, we can also obtain additional estimation of joint positions and orientations using the human body keypoint detection libraries OpenPose [15] and Blazepose [13]. Moreover, as the Vicon system is considered the best system for precision, for Group3, we also recorded with MoCap using the Vicon system. For synchronization purpose, the two systems were activated simultaneously.

3 METHODOLOGY

3.1 Gaussian Mixture Model

Gaussian mixture model (GMM) belongs to a group of probabilistic models and is used for representing data with a mixture of Gaussian probability density functions.

As in [11], we encode the movement point positions as a Gaussian Mixture Model (GMM): $\theta = [t, x]$, where t is the timestamp and x the joints positions.

$$p(\theta) = \sum_{i=1}^{K} \phi_i \mathcal{N}(\mu_i, \Sigma_i) \tag{1}$$

, where the i^{th} vector component is characterized by normal distributions with weights ϕ_i, means μ_i, and covariance matrices Σ_i. Each Gaussian of the mixture is thus defined by:

$$\mu_i = \begin{bmatrix} \mu_i^t \\ \mu_i^x \end{bmatrix}, \Sigma_i = \begin{bmatrix} \Sigma_i^t & \Sigma_i^{xt} \\ \Sigma_i^{xt} & \Sigma_i^x \end{bmatrix} \tag{2}$$

, where the indices ^{t} and ^{x} refer to respectively time and position.

The parameters ϕ_i, μ_i, Σ_i are learned by Expectation-Maximisation (EM) from the skeleton data of the movements captured by the Kinect or estimated with OpenPose or BlazePose.

For motion assessment, GMM is trained only on correct performances. In order to evaluate an observed sequence X during tests, we consider the negative log-likelihood that the given sequence X has been generated by the learned GMM. Correct motion sequences are expected to result in lower negative log-likelihood, in comparison to incorrect motion sequences. Hence, in order to detect incorrect motion sequences, we consider all motion sequences whose negative log-likelihood is higher than a threshold.

3.2 GMM in Riemannian manifold

While joint positions are naturally viewed in 3D Euclidean space, quaternions that we obtain from Kinect data having both position and orientation can be represented as elements of the 3-sphere S^3, which is a 3 dimensional Riemannian manifold. Such space is not linear as Euclidean, so the calculation of the mean and the covariance is not quite possible. However, we can consider tangent spaces as a linear approximation and map a point from the manifold to the tangent space and vice versa.

Table 1: F1 scores on Kinect trained data

Ex. type / train gr	group1A	group2A	group3
CTK	0.57	0.92	0.95
ELK	0.25	0.53	0.34
RTK	0.48	0.88	0.62
	0.43	0.78	0.64

Table 2: F1 scores on OpenPose trained data

Ex. type / train gr	group1A	group2A	group3
CTK	0.4	0.94	0.56
ELK	0.25	0.67	0.51
RTK	0.45	0.97	0.63
	0.37	0.86	0.57

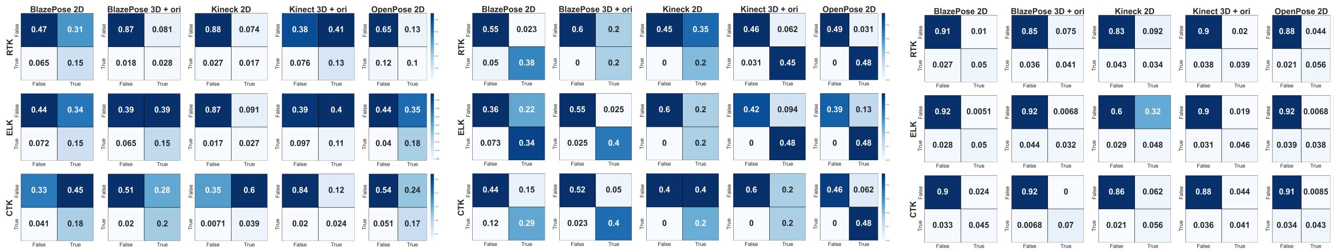
4 RESULTS AND DISCUSSION

We compare the ability of the two baselines to detect incorrect motion sequences, while only correct demonstrations are available during training. We evaluate our methods for Kinect, OpenPose and BlazePose data on three rehabilitation exercises for three different groups of performances, as explained above. In other words, we trained a different GMM model for each combination of data type used – Kinect / BlazePose (2D positions and 3D positions + orientations) and OpenPose (2D positions) – exercise type – CTK, ELK, RTK – and performing group –1A, 2A, and 3. To alleviate the issues caused by class imbalance and few data, for each such combination, we performed 4-fold cross-validation, thus we split correct performances into 4 groups and trained four GMMs, each time leaving a different fold for validation and training on the remaining three folds.

In order to evaluate the performance of GMMs, we compute the F1-score from correct and incorrect detections. We evaluate such a measure for varying thresholds from minimum to maximum value recorded for the negative log-likelihood of GMM baseline. We calculate negative log-likelihood values for all validation data (all incorrect performances for that combination and one fold of four of the correct performances, the one left for validation) and find the threshold to be used for classification by calculating F1 score for every possible threshold and choose the threshold with the biggest F1 score. Then, we average the obtained F1 scores for four 4 folds to obtain F1-score for one combination. These scores can be seen in Tables 1 and 2.

By comparing the reported results, we can see some differences in F1 and accuracy scores for one concrete combination of exercise type-performing group, but the overall scores are comparable. Differences in specific cases can be better understood by looking at the confusion matrices which are shown in Figures 4a, 4b and 5c, where we report the confusion matrices per group and exercise for each skeleton data, using either 2D positions with GMM or 3D positions and orientations with GMM in Riemannian manifold. They show that the confusion matrices are quite similar, thus that the model makes similar mistakes, for all the skeleton data analyzed. Thus, we can hypothesize that OpenPose, BlazePose or Kinect data are equivalent for automatic rehabilitation motion evaluation.

Marusic, A., Nguyen, S. M., and Tapus, A. (2023). Evaluating Kinect, OpenPose and BlazePose for Human Body Movement Analysis on a Low Back Pain Physical Rehabilitation Dataset. Companion of the 2023 ACM/IEEE International Conference on Human-Robot Interaction(587–591). Association for Computing Machinery.

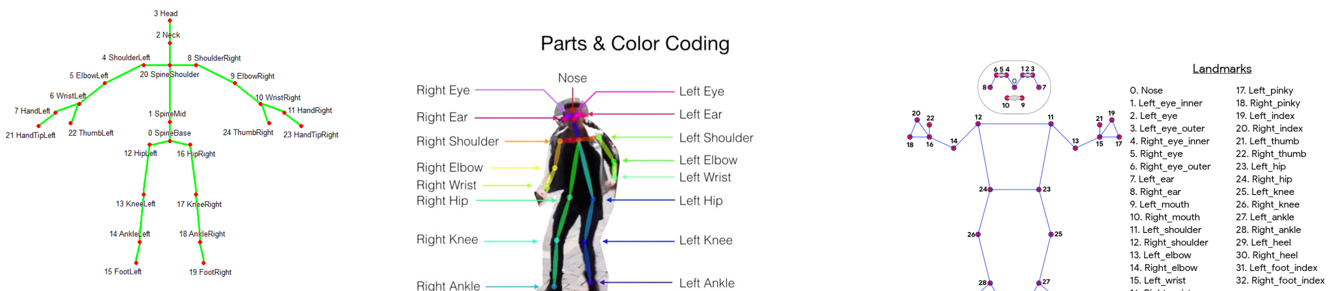


(a) Group 1A

(b) Group 2A

(c) Group 3

Figure 4: Confusion matrices for each performing group. Each row represents a data type (BlazePose 2D joint position, BlazePose 3D position + orientation, Kinect 2D, Kinect 3D + orientation, OpenPose). Each column represents an exercise (CTK, ELK, RTK).



(a) Kinect v2 - 3D positions and orientations of each joint of Kinect skeleton

(b) OpenPose - 2D positions of each joint of OpenPose skeleton

(c) BlazePose - 3D positions of each joint of BlazePose skeleton

Figure 5: Skeletons

For group3, exercise CTK, comparing Kinect and OpenPose data, we can see that the main problem is the quite big number of incorrect exercises. Therefore, even small differences in the number of correct performances classified wrongly make big differences in F1 scores between the two baselines.

On the other hand, by comparing Kinect and OpenPose data for exercise CTK for group 1A, we can see a slightly bigger number of incorrect exercises recognized as correct in OpenPose GMM. One possible explanation to this is the difference in calculated F1 scores across the 4 folds. In this specific case, the minimum F1 score was 0.35 while the maximum score was 0.53.

5 CONCLUSIONS

We identified the need for automatic coaching of physical rehabilitation exercises and the importance of quality motion assessment. We investigated whether a simple RGB camera has comparable results with Kinect by utilizing deep learning for pose estimation.

We evaluated and compared a baseline motion analysis algorithm, a Gaussian Mixture Model, on Kinect and OpenPose / BlazePose data, for the task of rehabilitation motion assessment. We demonstrated that, on average, results obtained through Kinect, OpenPose and BlazePose data were quite comparable, which shows that simple RGB cameras have a potential to be used as the main sensor for collecting movement data. As RGB cameras are easily accessible for laymen through any computer webcam or smartphone, and are often cheaper, we can expect an increasing use of OpenPose, Blazepose and similar pose estimation algorithms.

Although the analysis algorithm was not the focus in this work, we could see that the targeted tasks are still quite challenging. In that sense, some more sophisticated methods, possibly with the usage of deep learning and / or graph neural networks, can be designed in order to better assess rehabilitation movements.

REFERENCES

[1] Miron A, Sadawi N, Waidah I, Hussain H, and Grosan C. 2021. IntelliRehabDS (IRDS)—A Dataset of Physical Rehabilitation Movements. *Data* 6, 5 (2021). https://doi.org/10.3390/data6050046

[2] Tapus A., Bandera A., Vazquez-Martin R., and Calderita L. V. 2018. Perceiving the person and their interactions with the others for social robotics - A review. *Pattern Recognition Letters* (2018). https://doi.org/10.1016/j.patrec.2018.03.006

[3] Wu A, March L, Zheng X, Huang J, Wang X, Zhao J, Blyth FM, Smith E, Buchbinder R, and Hoy D. 2020. Global low back pain prevalence and years lived with disability from 1990 to 2017: estimates from the Global Burden of Disease Study 2017. *Ann Transl Med.* 8, 6 (Jan. 2020), 299. https://doi.org/10.21037/atm.2020.02.175

[4] Agathe Blanchard, Sao Mai Nguyen, Maxime Devanne, Mathieu Simonnet, Myriam Le Goff-Pronost, and Olivier Rémy-Néris. 2022. Technical Feasibility of Supervision of Stretching Exercises by a Humanoid Robot Coach for Chronic Low Back Pain: The R-COOL Randomized Trial. *BioMed Research International* 2022 (mar 2022), 1–10.

[5] Glowinski D, Dael N, Camurri A, Volpe G, Mortillaro M, and Scherer K. 2011. Toward a Minimal Representation of Affective Gestures. *IEEE Transactions on Affective Computing* 2, 2 (2011), 106–118. https://doi.org/10.1109/T-AFFC.2011.7

[6] Kulić D, Venture G, and Nakamura Y. 2009. Detecting changes in motion characteristics during sports training. *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (2009), 4011–4014.

[7] Maxime Devanne and Sao Mai Nguyen. 2017. Multi-level motion analysis for physical exercises assessment in kinaesthetic rehabilitation. In *2017 IEEE-RAS 17th International Conference on Humanoid Robotics (Humanoids)*. 529–534. https://doi.org/10.1109/HUMANOIDS.2017.8246923

[8] Maxime Devanne, Sao Mai Nguyen, Olivier Remy-Neris, Béatrice Le Gales-Garnett, Gilles Kermarrec, and André Thepaut. 2018. A Co-design Approach

for a Rehabilitation Robot Coach for Physical Rehabilitation Based on the Error Classification of Motion Errors. In *IEEE International Conference on Robotic Computing (IRC)*. 352–357.

[9] Aggarwal J.K. and Ryoo M.S. 2011. Human Activity Analysis: A Review. *ACM Comput. Surv.* 43, 3, Article 16 (apr 2011), 43 pages. https://doi.org/10.1145/1922649.1922653

[10] Matthieu L. 2014. *Poppy: open-source, 3D printed and fully-modular robotic platform for science, art and education.* Theses. Université de Bordeaux.

[11] Nguyen S M and Tanguy P. 2016. Cognitive architecture of a humanoid robot for coaching physical exercises in kinaesthetic rehabilitation. In *International Workshop on Cognitive Robotics*.

[12] Sao Mai Nguyen, Maxime Devanne, Olivier Remy-Neris, Mathieu Lempereur, and Andre Thepaut. 2024. A Medical Low-Back Pain Physical Rehabilitation Dataset for Human Body Movement Analysis. In *International Joint Conference on Neural Networks*.

[13] Bazarevsky V, Grishchenko I, Raveendran K, Zhu T, Zhang F, and Grundmann M. 2020. BlazePose: On-device Real-time Body Pose tracking. *CoRR* abs/2006.10204 (2020).

[14] Liao Y, Vakanski A, Xian M, Paul D, and Baker R. 2020. A review of computational approaches for evaluation of rehabilitation exercises. *Computers in Biology and Medicine* 119 (2020), 103687.

[15] Cao Z, Hidalgo Martinez G, Simon T, Wei S, and Sheikh Y. A. 2019. OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2019).