Filling gaps in motion capture data: a comparison of three statistical methods

Filling gaps in motion capture data: A comparison of three statistical methods

Gabriele Romano[1][0009−0007−6666−1724], Selene Perazzini[2][0000−0003−0292−9975], and Giorgio Gnecco[3][0000−0002−5427−4328]

[1] Casa Paganini - InfoMus, DIBRIS, University of Genova, Italy
gabriele.romano@edu.unige.it
[2] Faculty of Economics and Management, Free University of Bozen-Bolzano, Italy
selene.perazzini@unibz.it
[3] AXES Research Unit, IMT - School for Advanced Studies Lucca, Italy
giorgio.gnecco@imtlucca.it

Abstract. This study proposes the application of three methods from the statistical literature for reconstructing the position of an occluded marker in Motion Capture (MoCap) data. Specifically, we investigate the use of Gaussian Process Regression (GPR), the Synthetic Control Method (SCM), and a Matrix Completion (MC) algorithm. Unlike traditional gap-filling techniques, which rely solely on the occluded marker's time series, these methods exploit relationships between observed and missing markers to improve reconstruction accuracy. To assess their effectiveness, we apply these methods to a MoCap dataset comprising 3656 frames of a 3D human body performing approximately 17 distinct movements. A nested k-fold cross-validation framework is implemented using a selection of markers covering all major body parts. Performance is evaluated using the Mean Absolute Error (MAE). Our findings indicate that all three methods achieve satisfactory reconstruction accuracy, although computational efficiency varies significantly. The MC algorithm produces the most accurate results while also being the fastest method, whereas SCM exhibits the lowest accuracy. Finally, the methods are applied to real occlusions in the available dataset.

Keywords: Motion capture · Gaussian process regression · Synthetic control method · Matrix completion.

1 Introduction

Motion Capture (MoCap) is a technique used to record the movement of objects or individuals and is widely adopted in various fields, including video games, virtual reality, and medicine [9, 10]. A typical MoCap system consists of synchronised optical cameras that track reflective markers attached to a performer's suit. Each camera captures 2D coordinates, which are then triangulated into 3D space, generating a cloud of unlabelled points per frame. To obtain a structured representation of movement, a labelling process assigns unique identifiers to these

points using spatial and temporal heuristics combined with a voting algorithm. The final output is a time-sequenced recording of labelled markers.

A major challenge in this process is marker occlusion, which occurs when a marker is obstructed from all or a sufficient number of cameras, causing a loss of track or ambiguity during triangulation. This can result in the marker either being misidentified or not being registered by the system at all. Occlusions may result from various factors, such as self-occlusion (when a body part blocks a marker) or interference from other markers. To address this issue, some gap-filling techniques have been proposed for reconstructing missing marker data. Among these, most commercial MoCap softwares rely on interpolation methods or Kalman filters (e.g., [4–7, 13]), which estimate missing data based on the observed positions of the occluded marker. While these approaches are computationally efficient and easy to implement, their accuracy deteriorates when occlusions occur at the beginning of a recording session, persist for extended periods, or involve complex marker movements [8]. Beyond these approaches, it is important to acknowledge the existence, also in commercial softwares, of methods that incorporate mechanical constraints between markers, such as rigid body relationships and rotational constraints. These methods exploit such constraints to fill gaps by utilizing the motion of other markers in the scene. This is particularly useful in cases where multiple markers are occluded, yet enough remain visible to allow the system to infer missing data based on the movement of the rigid body as a whole, preserving the structural integrity of the captured motion. However, our analysis focus exclusively on uninformed methods.

This study examines the potential of three statistical approaches for reconstructing missing marker data: Gaussian Process Regression (GPR) [11], the Synthetic Control Method (SCM) [1], and Matrix Completion (MC) [3]. Unlike traditional interpolation and Kalman filtering, which rely solely on the occluded marker's time series, these methods exploit the relationships between observed and missing markers to enhance reconstruction accuracy. To assess their performance, we analyse a MoCap dataset comprising 3657 frames of a 3D human body performing approximately 17 distinct movements. A nested k-fold cross-validation is conducted on a subset of fully observed markers, selected to represent all key body parts (head, arms, hands, trunk, legs, and feet). The methods are evaluated based on their Mean Absolute Error (MAE). The validation results indicate that, while all three methods achieve satisfactory reconstruction accuracy, MC consistently yields the most precise estimates with the lowest variance, whereas GPR, despite comparable accuracy, incurs significantly higher computational costs. SCM, on the other hand, is the fastest method but also results in the least accurate predictions. Finally, the approaches are applied to real occlusions in our dataset.

This paper is organized as follows: Sect. 2 presents the data, Sect. 3 introduces the three algorithms, Sect. 4 illustrates the cross-validation and presents the related results. Sect. 5 applies the methods to the real occlusions in the dataset. At last, Sect. 6 concludes and discusses some future developments.

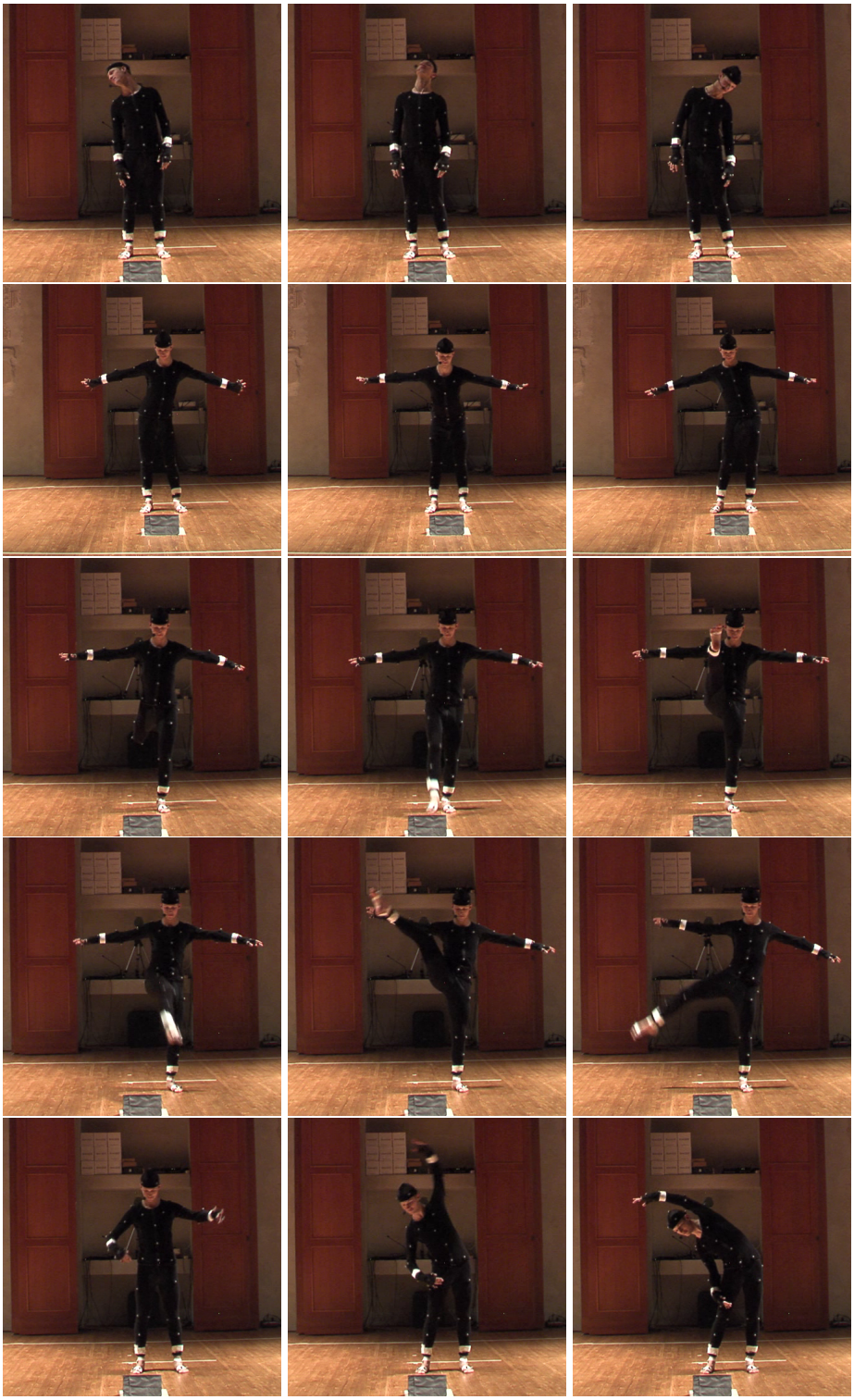

Fig. 1: Sequential snapshots from the analysed MoCap dataset, illustrating a subset of different movements performed by the subject.

2 Data

In the available MoCap dataset[4], the recording captures a professional dancer performing a sequence of dynamic warm-up exercises in a smooth and continuous manner. In total, the sequence comprises approximately 17 distinct movements, each lasting around two seconds. A selection of these movements is illustrated in Fig. 1. The dataset was acquired using the Qualisys MoCap system[5], which tracks marker trajectories with submillimetre precision, enabling accurate measurement of human motion. The recording consists of 3657 frames (captured at 100 frames per second), with each frame storing the spatial coordinates (x, y, z) of 64 reflective markers. Structurally, this dataset is represented as a 3657×192 matrix, where each row corresponds to a video frame, and each column represents a marker coordinate.

To refine the dataset, the final frame - containing 46 occluded markers - is removed, resulting in 3656 frames. However, occlusions persist throughout the recording, leaving only 39 fully observed markers. To evaluate the performance of the three statistical reconstruction methods, a subset of 10 markers is selected for cross-validation[6], ensuring coverage of all major body parts. A graphical representation of the markers, categorised by type (partially occluded, fully observed, and selected fully observed markers for validation), is shown in Figure 2. Moreover, the spatial coordinates of the markers are normalised so that each axis of the recording space ranges from 0 to 1 after normalisation.

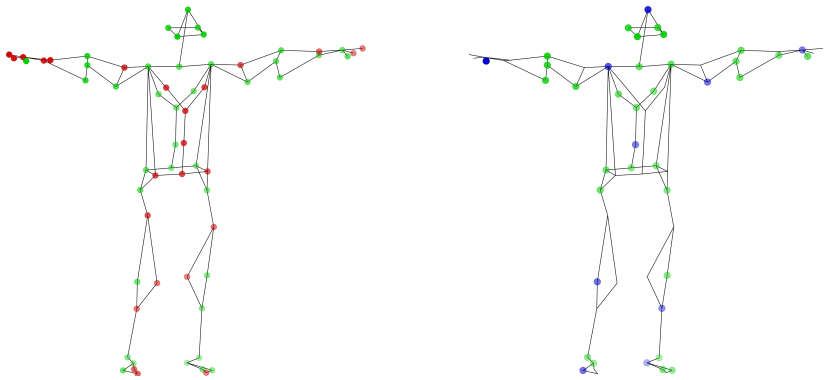

Fig. 2: Visualization of the 64 body markers recorded in the MoCap video. Left: markers without (green) and with (red) occlusions. Right: Markers selected for cross-validation (blue), which form a subset of the non-occluded markers (the remaining non-occluded markers are highlighted in green).

[4] Details are omitted in the first review round, due to the double-blind review process.
[5] https://www.qualisys.com/.
[6] Details on the specific cross-validation process are reported in Sect. 4.

3 Methods

This section provides a brief overview of the three methods proposed for reconstructing the position of an occluded marker: GPR, SCM, and an MC algorithm. Among existing MC techniques, we adopt the nuclear norm minimization approach proposed by [3], as it efficiently addresses non-random missing data, integrates fixed effects to improve estimation accuracy, and overcomes the computational limitations of fixed-rank methods. Moreover, while SCM and this form of MC were originally developed for counterfactual analysis, we show that they can be effectively adapted to the problem of missing marker position reconstruction. As for the implementation of the methods, the `sklearn.gaussian_process` module in `Python` was used for GPR, the `Synth` package in `R` was employed for SCM, and the `MCPanel` package in `R`[7] was used for MC.

Gaussian Process Regression (GPR) [11]. GPR is a Bayesian method that models a function $f(\boldsymbol{x})$ as a Gaussian Process (GP) characterized by a mean function $m(\boldsymbol{x})$ and a covariance function (kernel) $k(\boldsymbol{x}, \boldsymbol{x}')$. Given training data $\{(\boldsymbol{x}_i, y_i)\}_{i=1}^n$, where $\boldsymbol{x}$ are predictor variables and y is the target (one specific choice of target among the three normalised coordinates of an occluded marker, that we aim to predict), GPR assumes:

$$f(\boldsymbol{x}) \sim \mathcal{GP}\left(m(\boldsymbol{x}), k(\boldsymbol{x}, \boldsymbol{x}')\right). \tag{1}$$

A posterior distribution is computed by conditioning on observed data, allowing for uncertainty quantification in predictions.

Synthetic Control Method (SCM) [1, 2]. SCM constructs a synthetic trajectory for an occluded marker as a weighted sum of observed marker trajectories. Let y_t be the normalised coordinate of interest for the occluded marker at time t, and let $\boldsymbol{x}_t$ be the corresponding column vector of predictors (normalised coordinates of non-occluded markers). SCM estimates:

$$\hat{y}_t = \boldsymbol{x}_t^\top \boldsymbol{w}^*. \tag{2}$$

The weight column vector $\boldsymbol{w}^*$ is selected in such a way as to minimize the difference between observed and estimated values by solving a suitable nested optimization problem.

Matrix Completion (MC) [3]. MC reconstruct missing data by assuming that the elements in the complete data matrix show some degree of dependence. Consider a partially observed data matrix $\boldsymbol{M}$, where the rows represent statistical units and the columns report the categories of a variable. In our framework, $\boldsymbol{M} \in \mathbb{R}^{(3N-2)\times T}$, where the T columns refer to as many frames, the first $3N-3$

[7] This package was made available by the authors of [3] via GitHub (https://github.com/susanathey/MCPanel) and was not available on CRAN at the time of writing this paper.

rows refer to the three normalised coordinates of $N-1$ non-occluded markers, and the last row (which is partially observed) refers to the normalised coordinate of interest for the occluded marker. Then, according to MC, missing entries are estimated using nuclear norm minimization:

$$\left(\hat{\boldsymbol{L}}, \hat{\boldsymbol{\Gamma}}, \hat{\boldsymbol{\Delta}}\right) = \underset{\boldsymbol{L},\boldsymbol{\Gamma},\boldsymbol{\Delta}}{\arg\min} \left\{ \frac{1}{|O|} ||P_O\left(\boldsymbol{M} - \boldsymbol{L} - \boldsymbol{\Gamma}\mathbb{1}_T^\top - \mathbb{1}_N^\top \boldsymbol{\Delta}^\top\right)||_F^2 + \lambda ||\boldsymbol{L}||_* \right\}, \quad (3)$$

where $P_O(\cdot)$ selects the subset of observed elements of $\boldsymbol{M}$, $\boldsymbol{L}$ estimates its low-rank component, $\boldsymbol{\Gamma}, \boldsymbol{\Delta}$ account for row and column fixed effects, and $\mathbb{1}_T$ and $\mathbb{1}_N^\top$ are column vectors made respectively of T and N ones. The nuclear norm regularization term $\lambda||\boldsymbol{L}||_*$ prevents overfitting and ensures numerical stability.

4 Validation

To evaluate the performance of the three methods described in Sect. 3 for reconstructing the position of a missing marker, we conduct a series of numerical experiments in which a single marker is artificially occluded for a sequence of consecutive frames. To ensure the robustness of the estimation results and provide a comprehensive comparison of the methods, we implement a nested k-fold cross-validation strategy that accounts for temporal dependencies in the data.

Each method requires the optimisation of a distinct set of hyperparameters to achieve accurate predictions. For GPR, the prior distribution of $f(\boldsymbol{x})$ must be specified. Considering the data characteristics and computational constraints, we assume $m(\boldsymbol{x}) = 0$ and model marker trajectories using a constant kernel combined with a rational quadratic kernel. The hyperparameters of these kernels are then optimized. In SCM, the hyperparameter to optimize is related to the weight matrix, which determines the optimal combination of observed marker trajectories for reconstruction, while for MC, it is the constant λ in the nuclear norm penalty term. To avoid issues of data leakage and overly optimistic performance estimates that could arise from using the same data for both tuning and training, we decided to adopt a dataset partitioning into training, validation, and test sets. The hyperparameters' tuning is performed by minimizing the Mean Absolute Error (MAE) on the validation set.

In light of this, the nested k-fold cross-validation employed consists of an outer loop (with k=17) for model evaluation and an inner loop (with k=5) for hyperparameter tuning, ensuring robust performance assessment across different segments of the dataset. The partitioning process is structured as follows:

1. The subject in our MoCap video performs approximately 17 continuous movements of similar duration. Accordingly, we divide the dataset into 17 folds for the outer loop. To balance fold sizes and ease computation, the last frame of the video is removed, yielding 3655 frames.
2. In each replication of the outer loop, one fold is designated as the test set, comprising 215 frames.

3. The remaining 3440 frames are further split into 5 inner folds for hyperparameter tuning. In each iteration of the inner loop, one fold serves as the validation set (688 frames), while the remaining four constitute the training set (2752 frames).

This process is repeated for each of the 10 markers selected for validation, resulting in 17 predictions per marker, each spanning 215 consecutive frames, for every (normalised) coordinate and method. Each coordinate of an occluded marker is predicted separately, using the three coordinates of the remaining $N - 1 = 38$ fully observed markers as predictors.

The cross-validation results are presented in Fig. 3, which displays boxplots of the MAE for each selected marker. No substantial differences emerged in the ability of the methods to predict the three spatial coordinates. Consequently, the results are reported collectively across all three dimensions. Moreover, it can

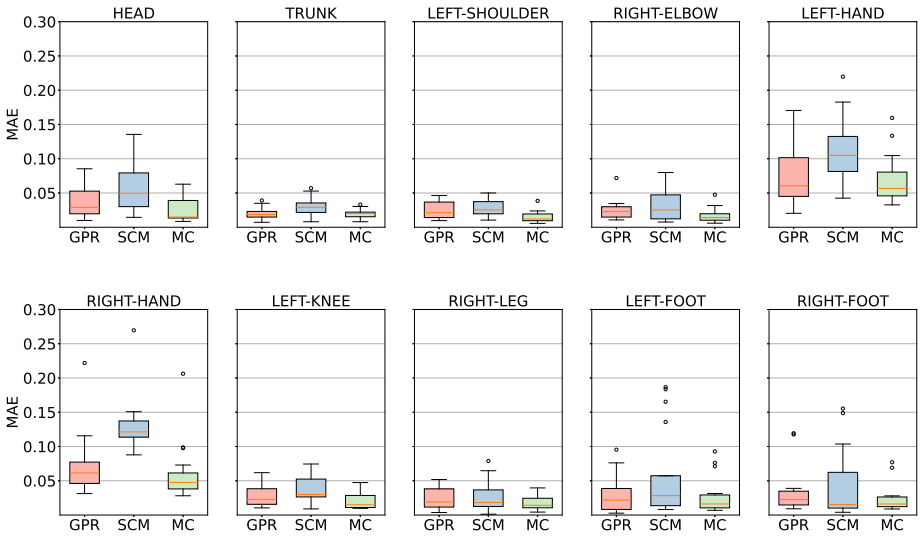

Fig. 3: Boxplots of the MAE for the 10 selected body markers.

be noticed that the performance is generally poorer for markers positioned on the hands, whereas for all other markers, the outcome is similar. This pattern is consistently observed across all three methods, indicating that predicting hand positions is inherently more challenging.

Overall, all three methods accurately estimate the marker positions along each axis. However, SCM consistently performs slightly worse than the other two, exhibiting a higher median MAE and greater variance. The performance of GPR and MC is comparable, although MC shows lower variance, making it a preferable choice. Moreover, MC is also a quite fast method, while GPR requires the highest computational time. Insights on the computational efficiency of each

method are provided in Table 1, which reports the average computation time (i,e., the average time for performing the whole nested cross-validation process) for predicting one spatial coordinate per marker and method. It emerges that MC is the fastest, followed by SCM, while GPR is significantly slower. This suggests that MC is not just preferable to GPR in terms of variance, but also in terms of the lower computational burden.

Table 1: Average computation time (in seconds) required to estimate a single spatial coordinate per marker and method. The computation times were measured on an Alienware Aurora R16 with an Intel Core i9-14900F 32-core processor and 64GB of RAM, parallelizing the 17 outer folds.

Marker	GPR	SCM	MC
HEAD	1013.48	67.99	28.61
LEFT-FOOT	1045.92	65.57	27.84
LEFT-HAND	1143.85	64.81	44.63
LEFT-KNEE	958.71	63.91	33.99
LEFT-SHOULDER	839.36	68.73	35.36
RIGHT-ELBOW	984.88	66.71	34.61
RIGHT-FOOT	1133.05	63.12	47.15
RIGHT-HAND	1115.73	65.85	52.87
RIGHT-LEG	932.38	61.51	59.35
TRUNK	1014.86	63.26	22.02

5 Application to real occlusions

Since the cross-validation analysis in Sect. 4 indicated that all three methods effectively reconstruct occluded markers, we apply them to real occlusions in our dataset. As in the cross-validation scenario, each coordinate of an occluded marker is reconstructed separately using the three coordinates of the fully observed markers (green points in Fig. 2, left) as predictors.

A graphical summary of the results is provided in Fig. 4, which showcases various occlusion scenarios to assess method performance under different conditions. In particular, we consider four representative cases: a single marker occluded for a short interval (row 1), a single marker occluded for a longer duration (row 2), multiple markers occluded over a short period (row 3), and multiple markers occluded over an extended interval (row 4). The figure clearly shows that the trajectories reconstructed using SCM (red lines) are consistently noisier compared to those reconstructed by the other two methods. In contrast, both GPR and MC tend to yield smoother, and thus more plausible, trajectories that align more closely with the natural dynamics of human motion. These observations are consistent with the cross-validation results.

Between GPR and MC, reconstructed trajectories are often remarkably similar - at times nearly indistinguishable. Nevertheless, in certain instances, GPR appears to capture the dynamics of human motion more realistically. This is particularly evident in the case of a prolonged occlusion of a foot marker during a rotational movement of the right leg (row 2), where the trajectory predicted

by MC lies somewhat between those of GPR and SCM, but appears less smooth and exhibits unnatural discontinuities. In contrast, GPR produces a continuous and physiologically plausible reconstruction that more accurately reflects the expected leg movement. To a lesser extent, similar differences emerge in the hand and arm movements in row 3 of the figure, where multiple finger markers are occluded for a short period. These findings underscore the importance of further investigations aimed at better understanding the capacity of the methods to generalise across movements of varying complexity and occlusion lengths. Nonetheless, it is important to emphasise that all reconstructed trajectories remain physically plausible. That is, for any given frame, each method generates predictions that are anatomically consistent with the visible body parts and represent feasible configurations within the natural range of human motion.

6 Conclusions and future developments

In this study, three methods originally developed in statistics have been applied to the reconstruction of missing markers. While the results are promising, this work constitutes a preliminary assessment of their potential, and further investigation is required. One key aspect that warrants closer examination is the duration of occlusions. Occlusions may persist for varying lengths of time, and the three algorithms may differ in their ability to predict short and long occlusions. Additionally, the methods may perform differently across various types of movement, and certain body parts may present greater challenges in marker reconstruction. These aspects merit further exploration.

A potential avenue for improving the accuracy of these approaches lies in the more precise selection of predictors for the occluded marker. This would necessitate a data pre-processing strategy (based, e.g., on clustering [12]) capable of effectively identifying informative markers to be incorporated into the specific algorithm. Such an approach could enhance predictive accuracy by reducing information redundancy while also improving computational efficiency. Furthermore, a comparison with established benchmark methods in the field would provide valuable insight into whether these techniques could serve as viable alternatives to the commonly used missing marker position reconstruction methods implemented in commercial MoCap software. As a further research direction, an important question is whether statistically significant differences can be identified through an appropriately chosen statistical test, taking into account issues such as the presence, for each method, of a partial overlap of training data in the construction of its associated 10 boxplots, shown in Fig. 3.

Acknowledgments. The work was supported in part by the PRIN 2022 project "MAHATMA" (CUP: D53D23008790006) and by the PRIN PNRR 2022 project "MOTUS" (CUP: D53D23017470001), both funded by the European Union – Next Generation EU program.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

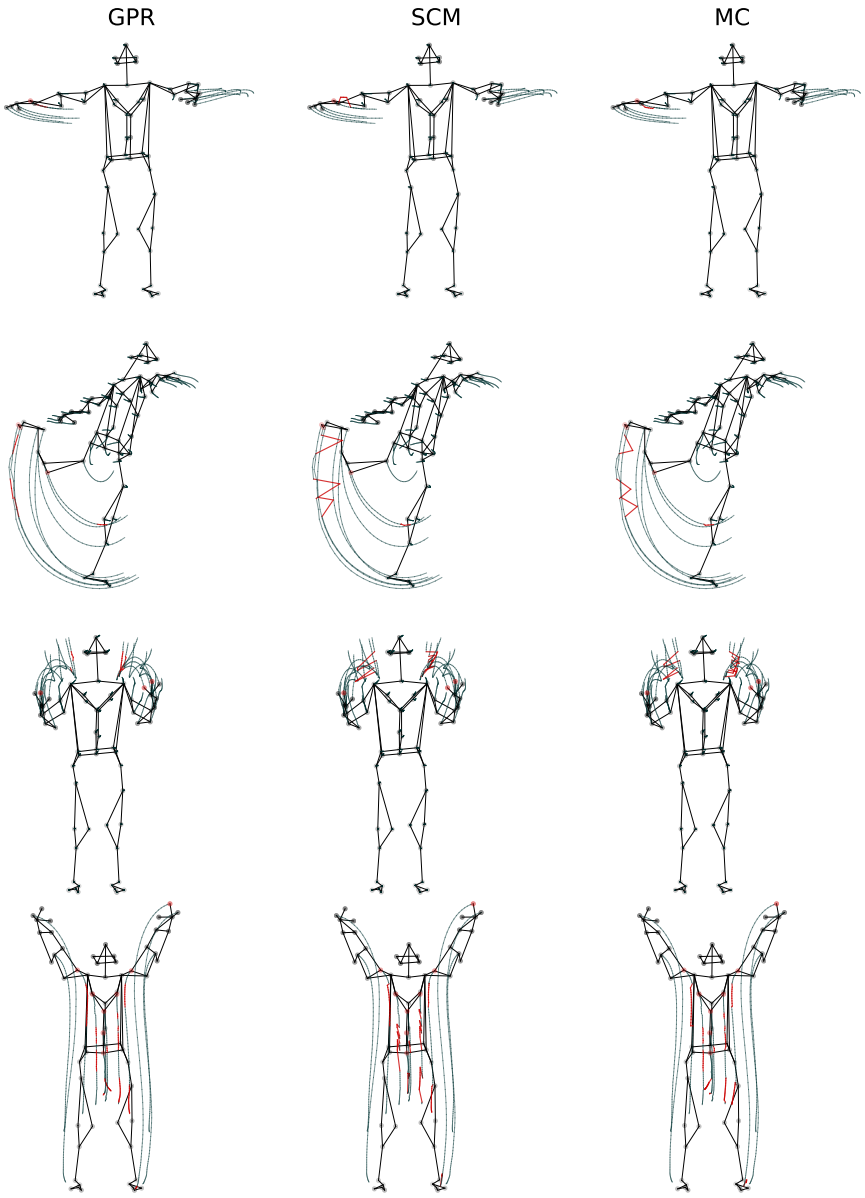

Fig. 4: Examples of 4 movements (one for each row) characterized by real occlusions in the MoCap dataset (where no MoCap ground truth is available), reconstructed (in red) by each method (one for each column). Each reconstructed trajectory is highlighted in red.

References

1. Abadie, A., Diamond, A., Hainmueller, J.: Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program. Journal of the American Statistical Association **105**, 493–505 (2010). https://doi.org/10.1198/jasa.2009.ap08746
2. Abadie, A., Gardeazabal, J.: The economic costs of conflict: A case study of the Basque country. American Economic Review **93**, 113–132 (2003). https://doi.org/10.1257/000282803321455188
3. Athey, S., Bayati, M., Doudchenko, N., Imbens, G., Khosravi, K.: Matrix completion methods for causal panel data models. Journal of the American Statistical Association **116**(536), 1716–1730 (2021). https://doi.org/10.1080/01621459.2021.1891924
4. Bonnet, V., Richard, V., Camomilla, V., Venture, G., Cappozzo, A., Dumas, R.: Joint kinematics estimation using a multi-body kinematics optimisation and an extended Kalman filter, and embedding a soft tissue artefact model. Journal of Biomechanics **62**, 148–155 (2017). https://doi.org//10.1016/j.jbiomech.2017.04.033
5. Burke, M., Lasenby, J.: Estimating missing marker positions using low dimensional Kalman smoothing. Journal of Biomechanics **49** (2016). https://doi.org/10.1016/j.jbiomech.2016.04.016
6. Cerveri, P., Pedotti, A., Ferrigno, G.: Robust recovery of human motion from video using Kalman filters and virtual humans. Human Movement Science **22**(3), 377–404 (2003). https://doi.org/10.1016/S0167-9457(03)00004-6
7. Halvorsen, K., Johnston, C., Back, W., Stokes, V., Lanshammar, H.: Tracking the motion of hidden segments using kinematic constraints and Kalman filtering. Journal of Biomechanical Engineering **130**(1), 011012 (2008). https://doi.org/10.1115/1.2838035
8. Liu, G., McMillan, L.: Estimation of missing markers in human motion capture. Visual Computer **22**(9-11), 721–728 (2006). https://doi.org/10.1007/s00371-006-0080-9
9. Moeslund, T.B., Granum, E.: A survey of computer vision-based human motion capture. Computer Vision and Image Understanding **81**(3), 231–268 (2001). https://doi.org/10.1006/cviu.2000.0897
10. Moeslund, T.B., Hilton, A., Krüger, V.: A survey of advances in vision-based human motion capture and analysis. Computer Vision and Image Understanding **104**(2), 90–126 (2006). https://doi.org//10.1016/j.cviu.2006.08.002
11. Rasmussen, C.E., Williams, C.K.I.: Gaussian Processes for Machine Learning. Adaptive Computation and Machine Learning, MIT Press (2006). https://doi.org//10.7551/mitpress/3206.001.0001
12. Romano, G., Rajeev Sabharwal, S., Gnecco, G., Camurri, A.: A computational framework for identifying salient moments in motion capture data. In: Nicosia, G., Ojha, V., Giesselbach, S., Pardalos, M.P., Umeton, R. (eds.) Machine Learning, Optimization, and Data Science. pp. 270–281. Springer Nature Switzerland, Cham (2025). https://doi.org/10.1007/978-3-031-82484-5_20
13. Senesh, M., Wolf, A.: Motion estimation using point cluster method and Kalman filter. Journal of Biomechanical Engineering **131**(5), 051008 (2009). https://doi.org/10.1115/1.3116153