A Pipeline for the Automatic Evaluation of Dental Surgery Gestures in Preclinical Training from Captured Motions

Mohamed Nail Hefied[1], Ludovic Hamon[1] and Sébastien George[1], Valériane Loison[2], Fabrice Pirolli[2], Serena Lopez[2] and Raphaëlle Crétin-Pirolli[2]

[1]*LIUM, Le Mans Université, France*

[2]*CREN, Le Mans Université, France*

mohamed_nail.hefied, ludovic.hamon, sebastien.george, valeriane.loison, fabrice.pirolli, raphaelle.pirolli @univ-lemans.fr,

Keywords: Automatic Gesture Evaluation, Motion Capture, Dental Surgery.

Abstract: This work in progress proposes an automatic evaluation pipeline for dental surgery gestures based on teacher's demonstrations and observation needs. This pipeline aims at supporting learning in preclinical situations for the first years of study in the dental school. It uses the Random Forest (RF) algorithm to train a model based on specific descriptors for each gesture component, that are designed to cover the evolution of the observation needs. The inputs are the captured motion parts whose labels are defined by the teachers with their own vocabulary, to represent expected or no-wanted geometrical or kinematic features. The overall evaluation (for example, weighted average of each component) and the component evaluation can be given to students to improve their postures and motor skills. A preliminary test correctly classifies a back correct posture and three main flaws ("Twisted Back and Bent Head", "Leaning Back", "Leaning Back and Bent Back") by the RF model, for the posture component. This approach is designed for the adaptation to the expert's evolving observation needs while minimizing the need for a heavy re-engineering process and enhancing the system acceptance.

1 INTRODUCTION

Training in dentistry begins with a preclinical period, dedicated to the learning of the most common procedures such as clinical examination, cavity preparation, tooth preparation for the crown placement, etc. It is also important to learn how to adopt the right postures to preserve the practitioner health and prevent pathologies, such as the development of MusculoSkeletal Disorders (MSD) (FDI, 2021).

During the preclinical period, students train on conventional simulators, also known as physical simulators or "phantom", consisting mainly of (i) a mannequin head (fig.3(a)) (ii) a model of jaws with artificial teeth (e.g., resin) that the students can insert into the phantom mouth (iii) and various instruments, including, mouth mirror, dental probe, rotative instruments, etc. In the dental school of Nantes University (France), a practical session typically includes, at least, twenty students. The assistance of teachers is often required by a student, making them unavailable to assist, assess and correct other students' gestures.

Alongside conventional simulators, there are virtual and haptic environments for dentistry training such as the HRV Virteasy Dental or NISSIN SIMONDOT system (Bandiaky et al., 2023). These simulators use force-feedback arms to replicate physical contacts of tools with the virtual teeth. The SIMTOCARE Dente training system includes a phantom's head on which augmented pedagogical feedback is provided. In both cases, those simulators primarily track the instrument movements through the haptic arm or motion sensors. It is therefore impossible to capture the user's body movements, which limits their ability to evaluate the full range of a student's technical gesture.

Nowadays, with motion capture (mocap) systems, one can record any motion-based activities to analyze and evaluate them and/or to build a dedicated Virtual Learning Environment (VLE) in which 3D avatars of the teachers and learners can be displayed in real time (Djadja et al., 2020; Le Naour et al., 2019). Mocap solutions also include those based on pose estimation in computer vision (e.g. OpenPose and MediaPipe). Movement data are often represented by a tree of joints (or skeleton) as shown in fig.3(e), each node

Hefied, M., Hamon, L., George, S., Loison, V., Pirolli, F., Lopez, S. and Crétin-Pirolli, R.
A Pipeline for the Automatic Evaluation of Dental Surgery Gestures in Preclinical Training from Captured Motions.
DOI: 10.5220/0012685000003693

In *Proceedings of the 16th International Conference on Computer Supported Education (CSEDU 2024) - Volume 1*, pages 420-427
ISBN: 978-989-758-697-2; ISSN: 2184-5026

containing a time series made of the 3D positions and orientations of the join over time.

The gesture learning can be considered according to three non-exclusive viewpoints: (a) the observation and the imitation of the expert gestures (Liu et al., 2020; Oagaz et al., 2022) (b) the learning of geometric, kinematic and dynamic features or (c) sequence of actions (focusing on reaching specific discrete states of manipulated objects regardless of the user's underlying movements)(Djadja et al., 2020). In this context, the pedagogical strategy can vary from one teacher to another for the same task. However, most existing current evaluation systems or VLE, neglect the motion-based evaluation process (neglect (a) and (b)) or the adaptation to the teacher's expertise that restricts the tool acceptance.

Consequently, the aim of this study is to propose a method and an operationalizable architecture able to: (a) automatically evaluate the dental surgery gestures using mocap systems, (b) integrate the teacher's expertise, in terms of gesture execution, to assist the students in their learning.

The main **contributions** of this work are the following ones:
- An analysis and decomposition of the dental technical gestures into evaluable motion-based components, making it possible to provide feedback to the learner and integrate the teacher's expertise thanks to each component.
- The proposal of an operationalizable approach for the automatic evaluation that addresses the dental surgery gesture at the component level and as a whole.
- The challenges and solutions to make this approach adaptable to the teacher's evolving observation and analysis needs, independent of the simulator type (conventional or virtual and haptic), independent of the motion capture system used as long as it provides skeletal data, and independent of the task to learn.

Although machine learning algorithms are used in this work, this study does not aim at contributing to the machine learning domain. AI is a technical solution to address our questions, not a goal in itself. Moreover, the automatic evaluation of the procedure (e.g., quality of the preparation, shape of the teeth, % of the removed dental cavity) are currently out of the scope of this work in progress.

The article is structured as follows. Section 2 reviews automatic gesture evaluation in the literature. Section 3 presents a survey on teacher practices and the potential advantages of an automatic gesture evaluation system to assist them and preclinical students. Section 4 breaks down dental surgery gestures. Section 5 describes the proposed system architecture. Section 6 depicts initial validation tests and discusses the results and the proposal, while section 7 gives the perspectives of this work.

2 RELATED WORKS

Many studies focus on specific aspects such as the posture of oral health professionals (Bhatia et al., 2020; Maurer-Grubinger et al., 2021; Pispero et al., 2021; FDI, 2021). However, to our current knowledge, it seems that there are no studies that fully address the automatic evaluation of gestures in dental surgery.

In other application domains, the automatic evaluation of gestures relies mainly on the analysis of movement data based on the teacher expertise. For instance, in the table tennis domain, the expert observation needs, related to forehand and backhand stroke gestures, were translated into metrics computed from movement data, while the gesture acceptability values were extracted from the pre-recorded demonstrations of the expert with a tolerance factor (Oagaz et al., 2022). Another study focused on the performances of novice salsa dancers compared to regular dancers (Senecal et al., 2020). Following the suggestions of dance experts, three criteria were proposed to represent the essential salsa skills (Rhythm, Guidance, and Style). The gesture was not studied as a whole but as a set of components, each of them associated with specific descriptors based on two popular motion analysis systems (MMF and LMA). A review of expressive motion descriptors, all based on kinematic, dynamic, and geometric features, was conducted (Larboulette and Gibet, 2015). However, other kinds of descriptors exist. Regarding postures, several studies used a score-type descriptor called the Rapid Upper Limb Assessment (RULA) score (Maurer-Grubinger et al., 2021; Bhatia et al., 2020; Manghisi et al., 2022). The RULA score is an ergonomic measure that evaluates the postural risk of the body during a task. It assigns a rating on a scale (1 safe to 7 dangerous) to a posture. The RULA score was adapted to the practice of oral health professionals, evaluating the postural risk during a therapeutic act for approximately 60 seconds (Maurer-Grubinger et al., 2021). In most of the cases, the descriptors are specifically chosen or designed for the task to learn, leading to significant engineering challenges if the observation needs or the task evolve.

An approach adapted to the evolution of the expert's needs while minimizing the reengineering process relies on the motion capture of the expert combined with spatial similarity techniques such as the Dynamic Time Warping (DTW). DTW aims at comparing the shape of two-time series without considering the temporal aspect. The lesser the DTW score is the closer the two series are. An acceptance threshold must be empirically defined. For instance, a VLE (Liu et al., 2020) used DTW to compare the Tai Chi movements of a learner with a virtual coach replaying the pre-recorded movements of the expert, while providing a similarity score. However, it is reasonable to question the pedagogical effectiveness of such a kind of score-based approach as it does not provide information about specific incorrect aspects of the gesture.

Another method consists in using supervised machine learning algorithms trained on motion-based data. In the context of home-based physical therapies, a study adopted a two-step machine learning classification approach that recognized the exercise among 10 types and then, evaluated whether the exercise is correctly executed or not (García-de Villa, 2022). They collected and used data from four IMUs placed on volunteers' limbs that performed each exercise series four times. Putting aside the complex processing chain of motion capture, the expert can train the model to evaluate new movements. However, the system's output is binary and does not provide relevant feedback to correct or enhance the gesture. The IANB gesture (Inferior Alveolar Nerve Block anesthesia) was evaluated (Sallaberry et al., 2022). The VIDA Odonto simulator collected the position and orientation of the syringe in the virtual environment. Several features were extracted from the collected data (e.g., mean jerk, penetration angle) and a comparison of the performances of different machine learning classification algorithms (Naive Bayes, Random Forests, Multi-Layer Perceptrons, and Support Vector Machine) and feature selection/fusion algorithms (ReliefF and PCA) was carried out. However, the output only discriminated between expert and novice levels without more information to guide the learners. In a study aiming at assisting the rehabilitation process of stroke patients (Weiss Cohen and Regazzoni, 2020), authors developed a system based on a leap motion as a hand-tracking device. The hand movements of the physiotherapist served to build a reference model. The gesture must be repeated 20-30 times. Joint angles were extracted and stored in a vector for each frame for each sample. A KNN algorithm allowed averaging the vectors for each frame of an exercise during the training phase. The system generated feedback for each finger separately, based on the angle difference, indicating the gap with the reference movement. The output was divided into four flexible segments that can be defined by users. An evaluation system based on ML is often designed as a proof of concept for the automatic evaluation performances. Despite, most of the existing systems limit the evaluation to an overall appreciation, the definition of output classes by the teachers can be relevant in terms of pedagogical feedback.

Systems only based on low-level descriptors are not adaptable to the evolution of tasks or observation needs. DTW and ML existing approaches can counterbalance this issue, but often overlook pedagogical feedback. Regarding dental surgery techniques, there is not one perfect motion i.e. various biomechanical approaches are viable as long as they meet the experts' criteria for each gesture aspect. Consequently, there is a challenge in designing a system adapted to the teachers' practices and their evolution, while providing feedback to the learner related to those criteria. The expectations of such a system are discussed in the next section.

3 EVALUATION EXPECTATIONS

To gather teaching practices, observation and analysis needs in relation to technical gestures and their evaluation, a qualitative survey was carried out using semi-directive interviews with eleven teachers from several disciplines (Prosthetics, Restorative Dentistry and Endodontics (DRE) and Pediatric Dentistry) at the dental school of Nantes University. The semi-directive interviews, lasting an average of thirty-five minutes, were conducted by a pair made of a computer scientist researcher and an educational researcher. The interview recordings were anonymized and the audio files transcribed. A thematic analysis of the verbatim, by manually coding the discourse segments, was carried out using the Nvivo software. The main topics of the interview and their results are exposed below.

Work Session and Formative Evaluation. A Practical Work (PW) session, dedicated to working on conventional simulators, generally includes about *"twenty" (ens1)* students supervised by a teacher, assisted either by monitors (students with a higher level of study) or operating alone. A formative evaluation, without grading, is carried out at each stage of the exercise by the teacher or monitors, based on their observations or at the student's request: *"It's to validate the steps as they are carried out" (ens3);"each time, we give them a little advice on what was done well and what must be improved" (ens2);*

High Demand on Teachers and Procedure Results vs. Gesture Evaluation. The demands placed on teachers and monitors are considerable, due to the size of the student groups: *"when you're managing 20 or 23 students, that's a lot, and you're not necessarily available at the right time" (ens1).* Therefor, assessment primarily focuses on the result of the dental surgery procedure, rather than the technical gestures performed by students: *"after a while, even for us, we're human, so we end up looking mainly at the clinical aspect, the final result (...) whereas the means to achieve it, is very important" (ens2).*

Dental Preparation vs. Gesture Concerns. A session can last between 1h30 and 3h00, and the number and duration of PW sessions are limited by the density of the required teaching, *"the problem is that if at a given session, the student hasn't assimilated all information (...) it's a bit lost, given that at the next session we'll move on to another exercise" (ens7).* In addition to the preparation, *"for them, the working position, ergonomics, are a secondary objective (...) and they may find it easier or quicker to bend their neck to see better and perform the gesture" (ens3)* that is not recommended to avoid MSD.

Interests of Automatic Evaluation of Technical Gestures. The interest of automatic evaluation of technical gestures is to (i) *"help because it's complicated to manage all the students" (ens2)* for whom *"you have to repeat over and over again (...) and who regularly forget their work position" (ens3)* (ii) calculate metrics in real or near-real time, which can be used to provide feedback to the student on the technical gesture (e.g.: Your back is bent, stand straight, lower your elbows...) *"if it's something to tell to students, it might be more educational for them, and especially for all those we don't see at a given moment"(ens9)* (iii) standardize practices further, and *"perhaps smooth out the level between different groups a little more, and ensure that certain transmissions of information are not teacher-dependent, because certain choices of instruments, for example, certain set-ups, certain working position will be a matter of habit, a matter of personal feeling. (...) we don't practice in the same way (...) as long as it remains within a framework where it's done in good conditions"(ens5).*

An automatic system for evaluating students' gestures can help teachers, who face high demands and repetitive gesture issues. This system allows students to focus on procedures while being reminded of the correct gestures. This is possible if the dental surgery gesture is formalized in a frame made of interpretable, operationnalisable and evaluable components, allowing the integration of teachers' observation needs, as discussed in the next section.

4 DENTAL SURGERY GESTURE

In addition to interviews with teachers, two visits were made during PW sessions at the dental school of Nantes University, to observe and perform a PW (i.e., preparing a tooth for the placement of a crown). Furthermore, an analysis of the two PW notebooks of the teachers, in Prosthetics, Restorative Dentistry and Endodontics was conducted, along with a review of the 2021 ergonomic recommendations for oral health professionals (FDI, 2021) published by the FDI (World Dental Federation). Based on the gathered pieces of information, the surgery dental gesture can be broken down as follows.

Posture. This component qualifies the body part configuration to adopt: (i) The natural curvatures of the spine must be respected (cervical lordosis, thoracic kyphosis, lumbar lordosis). The forward body must not lean (bust/leg angle $\geq 90°$). No excessive bending or twisting of the spine (including back and head) must be observed. The head slightly tilts forward. (ii) Arms can be at rest or almost alongside the torso (20° between vertical and arms). There is no abduction of the shoulders. The practitioner's elbows are close to the body and do not protrude. Forearms are in front of the body (elbow angle 60°). Wrists held in a neutral and straight position. (iii) One must observe legs apart and lower legs vertical (knee angle 90° to 100° degrees). Feet must be flat on the floor.

Sitting Orientation. This component represents the practitioners' seated position around the patient's head, according to their dominant arm. The space occupied by a right-handed person must be between 9 and 12 o'clock (12 and 3 o'clock for left-handed one). In this interval, and depending on the tooth to be treated, the practitioners must opt for a positioning that enable them to better see the tooth.

Instrument Holding and Fulcrum (Finger Rest). Errors in holding rotative instruments are recurrent, and difficult to detect in a PW context, requiring the teacher to be close to each student's workstation.

The instrument should be held like a pen by three fingers (thumb, index and middle fingers) close to the head of the instrument, to control the pressure applied to the tooth. The other two fingers are positioned as close as possible to the preparation, ideally on the same working arch, acting as a fulcrum on the tooth or gum. The objective is not working with a floating hand but following the patient's arch to have an accurate motion, reduce muscular load and fatigue, and avoid injuring the patient.

Asepsis. In addition to complying with general asepsis guidelines, such as wearing goggles, masks, the gloves of the practitioner must not touch anything other than the patient's oral cavity (e.g., tooth, arch, gum), and the instruments placed on the operating field. The goal is to monitor parasitic movements e.g., scratching one's nose or head, leaving a free hand on

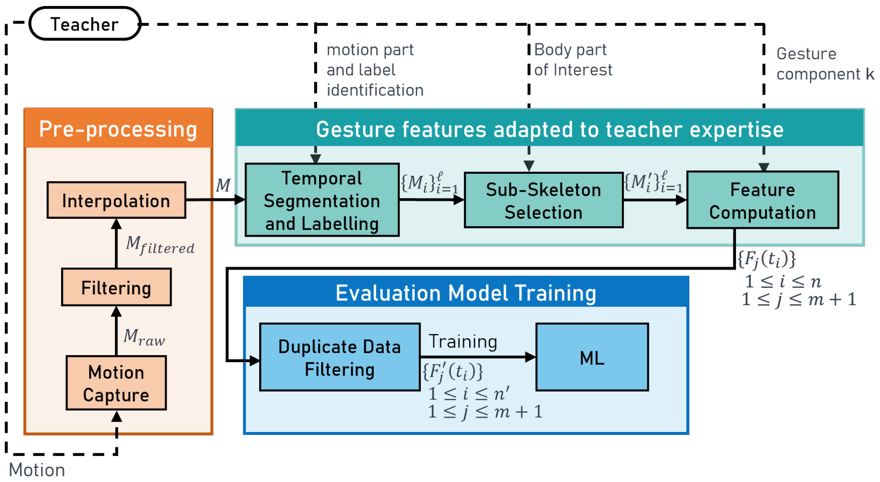

Figure 1: Training phase.

their own pants, etc., This kind of gestures leads to hygiene faults.

All the previous four components were validated by the three teachers implied in the visits. Despite the given specific measurements and values provided by the aforementioned pedagogical documents, the automatic evaluation system should encompass all gesture aspects, while being adaptable to changes and tolerance (e.g. is 91°, 92°, etc., an acceptable value for the bust/leg angle ?). Consequently, the next section outlines a pipeline based on gesture components, analyzed each with descriptors able to integrate the teachers' expertise from their demonstrations.

5 PROPOSED EVALUATION PIPELINE & METHOD

This section outlines the proposed evaluation pipeline, which includes training individual Machine Learning (ML) models for each gesture component, identified in the previous section, to conduct continuous learner assessments.

5.1 Training Phase

Fig.1 illustrates the initial phase of the ML model training, applicable for each gesture component, and using generic descriptors computed from labeled teacher demonstrations. Therefore, for a given component, this phase begins by asking the expert to provide good and/or bad demonstrations.

Motion Capture and Filtering. The expert's movements are captured using a motion capture system such the Qualysis infrared system. M_{raw} is the raw motion data structure. This raw data can be noisy, containing inconsistent or missing values, and must therefore be manually filtered (linear, polynomial, Savitzky Golay, relation filters, etc.) to obtain a clean motion $M_{filtred}$.

Interpolation. An interpolation process will generate the motion M with the desired frequency or frame number, as a mocap system with a high frequency can generate too much data than the system can handle in a reasonable time.

Temporal Segmentation and Labeling. From M, the experts (or teachers) identify (non-)acceptable sequences $\{M_i\}_{i=1}^{\ell}$ with ℓ being the number of labels. They must visualize the 3D motions and give the corresponding time periods and labels. The teacher may define and identify as many (non-)acceptable motion parts as they want in the way they want (e.g., correct, incorrect, almost correct, bending back, head leaning, weird shoulder position, etc.).

Sub-Skeleton Selection. From the complete joint tree, the teacher selects the branches (succession of joints) representing the body part of interest for a gesture component. This module returns the $\{M_i'\}_{i=1}^{\ell}$ labeled motion spatially trimmed to the desired set of joints.

Feature Computation. This module computes predefined features (or descriptors) for a given gesture component. The challenge here is to find an appropriate set of features adapted to the evolution of the teaching practices, i.e., any expert could integrate new gestures to identify without changing the descriptors for the component. The proposed features are the following ones:

- (i) Posture and instrument holding: the joint orientation (quaternion) from a movement expressed in a local coordinate system, plus normalized directional

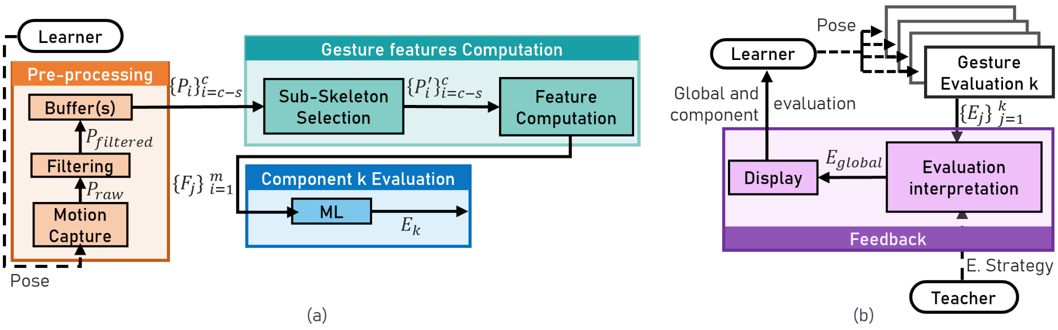

Figure 2: (a) Evaluation phase per gesture component (b) Evaluation feedback.

vectors from the root joint to each of the other joints computed from a movement expressed in a global coordinate system(fig. 3(e)).
- (ii) Sitting orientation: the angle between a straight line connecting the root joint to a reference point on the operating area, and a fixed reference line (e.g., 12 o'clock).
- (iii) Asepsis: distances between hand and wrist joints to other body joints.
The fulcrum is currently being studied and features will be proposed in the future. The output $\{F_j(t_i)\}$ of this module is a data table with n row (frame) and $m+1$ column (time series of features and a label column).
Duplicate Data Filtering. This module parses $\{F_j(t_i)\}$ with a sliding window to compare rows and filter out duplicates that do not meet a threshold to return $\{F_j'(t_i)\}$ with $n' < n$. The objective is to only keep distinctive necessary samples for the ML training process.
ML. The machine learning module will correlate the training samples to their expected labels defined by the teacher. The chosen algorithm is the Random Forest (RF). This algorithm is non-dependent to a distance function as heterogeneous component features can be considered in different kinds of distance function (Euclidean, spherical, geodesic, etc.). Decision trees generated by a Random Forest identify the most informative data divisions by maximizing information gain or minimize entropy. This algorithm is also known to perform well with few data as one cannot ask a teacher to make many demonstrations.

5.2 Evaluation Phase

Now that the ML model is trained with the teacher's (non)acceptable gesture demonstrations, it can be used to evaluate the learner's gestures. Figure 2(a) describes the pipeline to evaluate a single gesture component based on a capture of the learner's pose P (i.e., the joint tree in a single frame) for (near) real-time evaluation. $\{P_i\}$ contains s poses (i.e. short motion part) stored by the buffer module for later computation of descriptors requiring several poses (e.g., speed), while $\{P_i'\}$ is the sub-skeleton trimmed according to the targeted subset of joints of interest defined by the teacher. Finally, the ML block infers a class for the gesture component.

Figure 2(b) shows the feedback information sent to the learner. This feedback is composed of two kinds of information. The first information is the inferred class corresponding to the teacher's label for each gesture component. The second information, E_{global} gathers $\{E_j\}$ predictions to deliver a global evaluation, defined by the teacher's evaluation strategy (e.g., a score based on the weighted average of each digitally transformed component). The representation method (textual information ? Dashboard ? More advanced visual artefacts) of this evaluation is not defined at this stage of this work.

6 PRELIMINARY TESTS & DISCUSSION

The posture component based on the respect of the natural curvatures of the back was implemented. To this end, an installation was setup with a phantom attached to a table and a stool positioned at 10:30 from the head of the mannequin (figure3(a)). Surrounding this setup are 6 Qualisys Miqus M3 infrared cameras to capture the movements of the expert (at 100 Hz), who is equipped with an upper-body marker set. The figure3(e) illustrates a tree of joints. The expert simulates a therapeutic act on a tooth located on the upper arch, maintaining an acceptable posture (1 min.) and unacceptable ones (2 min.). The table 1 depicts the posture classes, the record sequence duration and the number of samples (i.e. number of frames) obtained after the duplicate data filtering. Figure3 shows each posture classes.

After training a RF model with 100 decision trees,

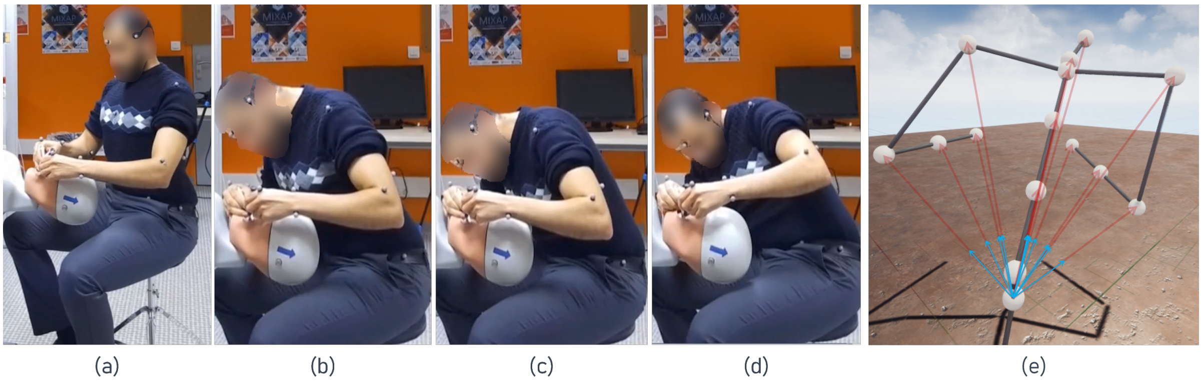

Figure 3: (a) Straight Back (b) Leaning Back (c) Leaning Back and Bent Back (d) Twisted Back and Bent Head (e) A skeleton from motion capture under Unreal Engine with directional vectors (red) and normalized ones (blue) starting from the pelvis joint, and pointing to the remaining ones.

based on 80% training 20% test split of the expert samples (all classes mixed), our architecture is able to perfectly recognized each good and bad posture (perfect accuracy score).

Table 1: Posture class, record sequence duration (seconds) and number of frames obtained after data temporal segmentation and duplicate data filtering.

Posture Class	**Time**	**Samples**
Straight Back	60	3273
Leaning Back	4	232
Leaning Back and Bent Back	4	182
Twisted Back and Bent Head	8	428

The above first tests offer encouraging results, but are not yet a proof of the system's validity and performance. Other experiments will be carried out with several teachers, their observation needs and dental students with different morphologies. Provided that the system proves its effectiveness, it offers the following advantages:

Adaptation to the Teachers' Needs. The teachers are actively involved in the proposed pipeline. Their expertise relies on several provided pieces of information: (non)acceptable motions, body parts of interest, labels linked to each aspect or skill related to the gesture to learn and the evaluation strategy combining each component. All these pieces of information does not impact the system reengineering given that: (a) the proposed set of components to analyze the dental gesture is valid (b) and the set of descriptors for each component is carefully chosen to cover different kinds of correct gestures and flaws for this component. In this way, the system becomes adaptable to each demonstrated gesture the teacher does or does not want to see.

Building Relevant Pedagogical Feedback. The system architecture handles the gesture as a set of evaluable components. For each of these components, a RF model is trained separately. This approach allows for an evaluation of the overall gesture without neglecting the assessment of each component representing the gesture aspect to acquire or to enhance. In addition, the evaluation uses the teacher's vocabulary thanks to the labels.

System Independencies. The proposed architecture is designed to be independent of any specific motion capture system (as long as it provides a skeleton made of the position and orientation of each joint), the used simulator (conventional or haptic), the task to learn and the pedagogical strategy (if based on the chosen valid components and their descriptors).

Nevertheless, the following limitations and remaining challenges must be considered:

MoCap Process. The process of obtaining a clean time series can be tedious. Indeed, some motion capture devices can be costly (infrared camera-based ones), give a good precision and require a heavy data pre-processing (marker re-identification, interpolation of lost data, etc.). Other systems are less costly and quicker to set up, but the signal quality is worse (depth camera, inertial units, etc.).

Proposed Architecture vs. ad-hoc Implementation. When the descriptor and range of acceptability are considered as trivial to implement (e.g., the sitting orientation around the phantom requires verifying an angle within a well-defined range), the question of the interest of an ML training process requiring the presence of a teacher, a capture session and a labeling process can be raised.

Learning Impact. The potential impacts of the system on learning must be considered with caution and must be tested. The observation needs are only formalized through the labels associated to the demonstrated motion parts. The system is not currently adapted to the formalization of the underlying knowledge related to the overall healthcare procedure.

7 CONCLUSION

This work proposes a pipeline for the automatic evaluation of dental surgery gestures. The aim of this system is to assist teachers and learners during practical sessions on simulators (conventional or virtual and haptic). The expected long-term impacts are related to the improvement of motor skills in preclinical situations, to prepare students for clinical ones, and avoid learning motions leading to MSD. This first step breaks down the gesture into components (posture, sitting orientation, holding the instrument, fulcrum, asepsis) and proposes generic descriptors for each component. The proposed approach consists in training random forest models for each component, whose inputs are the generic descriptors computed from the teacher's labeled and captured motions. Each label is defined by teachers to integrate the observation needs with their own vocabulary. The trained RF model can be used to analyse the learners' gestures by giving the class label for each gesture component. This architecture tends to tackle the challenges linked to the evaluation of the often neglected geometric and kinematic aspects of the dental gesture in the existing systems, while avoiding a heavy reengineering process in case of the evolution the learning situation. This work will continue through an experiment with a dual objective: (a) validating the pipeline in terms of evaluation performances with teachers and (b), evaluating the impact of the evaluation on students during practical sessions.

ACKNOWLEDGEMENTS

The authors would like to thank the support given by the French Research National Agency in funding of the ANR PRCE EVAGO project (ANR-21-CE38-0010).

REFERENCES

Bandiaky, O. N., Lopez, S., Hamon, L., Clouet, R., Soueidan, A., and Le Guehennec, L. (2023). Impact of haptic simulators in preclinical dental education: A systematic review. *Journal of Dental Education*, n/a(n/a). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/jdd.13426.

Bhatia, V., Randhawa, J. S., Jain, A., and Grover, V. (2020). Comparative analysis of imaging and novel markerless approach for measurement of postural parameters in dental seating tasks. *Measurement and Control*, 53(7-8):1059–1069.

Djadja, D., Hamon, L., and George, S. (2020). Design of a Motion-based Evaluation Process in Any Unity 3D Simulation for Human Learning:. In *Proceedings of the 15th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*, pages 137–148, Valletta, Malta. SCITEPRESS - Science and Technology Publications.

FDI (2021). Ergonomics and posture guidelines for oral health professionals.

García-de Villa, S. (2022). Simultaneous exercise recognition and evaluation in prescribed routines: Approach to virtual coaches. *Expert Systems With Applications*. titleTranslation:.

Larboulette, C. and Gibet, S. (2015). A review of computable expressive descriptors of human motion. In *Proceedings of the 2nd International Workshop on Movement and Computing*, pages 21–28, Vancouver British Columbia Canada. ACM.

Le Naour, T., Hamon, L., and Bresciani, J.-P. (2019). Superimposing 3D Virtual Self + Expert Modeling for Motor Learning: Application to the Throw in American Football. *Frontiers in ICT*, 6:16.

Liu, J., Zheng, Y., Wang, K., Bian, Y., Gai, W., and Gao, D. (2020). A Real-time Interactive Tai Chi Learning System Based on VR and Motion Capture Technology. *Procedia Computer Science*, 174:712–719.

Manghisi, V. M., Evangelista, A., and Uva, A. E. (2022). A Virtual Reality Approach for Assisting Sustainable Human-Centered Ergonomic Design: The ErgoVR tool. *Procedia Computer Science*, 200:1338–1346.

Maurer-Grubinger, C., Holzgreve, F., Fraeulin, L., Betz, W., Erbe, C., Brueggmann, D., Wanke, E. M., Nienhaus, A., Groneberg, D. A., and Ohlendorf, D. (2021). Combining Ergonomic Risk Assessment (RULA) with Inertial Motion Capture Technology in Dentistry—Using the Benefits from Two Worlds. *Sensors*, 21(12):4077. Number: 12 Publisher: Multidisciplinary Digital Publishing Institute.

Oagaz, H., Schoun, B., and Choi, M.-H. (2022). Real-time posture feedback for effective motor learning in table tennis in virtual reality. *International Journal of Human-Computer Studies*, 158:102731.

Pispero, A., Marcon, M., Ghezzi, C., Massironi, D., Varoni, E. M., Tubaro, S., and Lodi, G. (2021). Posture Assessment in Dentistry for Different Visual Aids Using 2D Markers. *Sensors*, 21(22):7717.

Sallaberry, L. H., Tori, R., and Nunes, F. L. S. (2022). Comparison of machine learning algorithms for automatic assessment of performance in a virtual reality dental simulator. In *Symposium on Virtual and Augmented Reality*, SVR'21, pages 14–23, New York, NY, USA. Association for Computing Machinery.

Senecal, S., Nijdam, N. A., Aristidou, A., and Magnenat-Thalmann, N. (2020). Salsa dance learning evaluation and motion analysis in gamified virtual reality environment. *Multimedia Tools and Applications*, 79(33):24621–24643.

Weiss Cohen, M. and Regazzoni, D. (2020). Hand rehabilitation assessment system using leap motion controller. *AI & SOCIETY*, 35(3):581–594.