VCoach: A Customizable Visualization and Analysis System for Video-based Running Coaching

Jingyuan Liu, Nazmus Saquib, Zhutian Chen, Rubaiat Habib Kazi, Li-Yi Wei, Hongbo Fu, and Chiew-Lan Tai

Abstract—Videos are accessible media for analyzing sports postures and providing feedback to athletes. Existing video-based coaching systems often present feedback on the correctness of poses by augmenting videos with visual markers either manually by a coach or automatically by computing key parameters from poses. However, previewing and augmenting videos limit the analysis and visualization of human poses due to the fixed viewpoints, which confine the observation of captured human movements and cause ambiguity in the augmented feedback. Besides, existing sport-specific systems with embedded bespoke pose attributes can hardly generalize to new attributes; directly overlaying two poses might not clearly visualize the key differences that viewers would like to pursue. To address these issues, we analyze and visualize human pose data with customizable viewpoints and attributes in the context of common biomechanics of running poses, such as joint angles and step distances. Based on existing literature and a formative study, we have designed and implemented a system, VCoach, to provide feedback on running poses for amateurs. VCoach provides automatic low-level comparisons of the running poses between a novice and an expert, and visualizes the pose differences as part-based 3D animations on a human model. Meanwhile, it retains the users’ controllability and customizability in high-level functionalities, such as navigating the viewpoint for previewing feedback and defining their own pose attributes through our interface. We conduct a user study to verify our design components and conduct expert interviews to evaluate the usefulness of the system.

Index Terms—Human Pose, Video Processing, Sports Data Analysis

1 INTRODUCTION

Running is a globally popular exercise, and many runners want to avoid injuries and improve their performance. Not everyone can have access to human coaches, and thus various online materials and mobile apps have emerged to provide guidance on achieving correct running forms. As with general sports training, an accessible means for novice sports players is to learn from pre-recorded performances of coaches or professional players by performing and comparing the same actions.

According to the taxonomy of comparison-based visualization [15], existing visualizations for human pose comparison include displaying related poses in two videos side-by-side (juxtaposition) [37, 47, 48], overlaying one pose onto another (superposition) [11], and augment-
When observing the actions in videos, the 3D human pose attributes worked closely with experts in Sports Science to identify its designs (see the toy example in Fig. 2). For sports coaching systems, such an ambiguity problem affects both the observation and the feedback. When observing the actions in videos, the 3D human pose attributes might be distorted due to perspective shortening and thus fail to reflect the actual biomechanical measurements. In visualization, the shapes of graphical annotation markers overlaid on videos are also subject to changing viewpoints, and are thus ambiguous in providing accurate corrective feedback to be perceived by amateur runners. To promote spatial awareness, prior studies have attempted to analyze reconstructed 3D poses [13], fuse videos in multiviews [46], and use situated AR [28] and immersive visualization [10, 24]. Thanks to the emerging methods in monocular human reconstruction in computer vision [9, 16], reconstructing 3D poses has become an effective and accessible solution for videos.

Besides the ambiguity problem, another consideration is the data attributes for comparison, which can be classified as parametric and non-parametric. Parametric pose features (e.g., knee angle) are sport-specific and pre-defined by domain experts [6]. The embedded bespoke knowledge makes sport-specific systems hard to scale and support users’ needs for individual customization. Alternatively, non-parametric comparison avoids embedding bespoke knowledge by comparing the transferred and overlaid human poses [11]. Novices would need to infer the corrective feedback based on their perceptions.

To address the above-mentioned issues, we aim to develop an interactive system to analyze and visualize differences in human biomechanical data. Our system, VCoach, provides intuitive and customizable corrective feedback for amateur runners. To achieve this goal, we worked closely with experts in Sports Science to identify its designs based on the coaching process in practice. As shown in Fig. 1, our system takes as input a sample user video from an amateur runner and an exemplar video from an expert runner, and automatically performs pose analysis tasks, such as reconstructing 3D poses from videos and computing pose differences. The differences are then visualized as short animations on a 3D human body model (Fig. 1(d)) to resemble the dynamic demonstration of human coaches in practice. To reduce the ambiguity of visualization, we propose to augment 3D visual markers onto the 3D body model instead of the video, such that users can either preview under our suggested viewpoints or manually navigate through viewpoints for better perception.

VCoach embeds pre-defined biomechanical attributes that are commonly used for analyzing running poses (e.g., leaning angle and foot landing position). To support the analysis of attributes users are interested in but not embedded in the system (e.g., vertical bend angle of knees and height of feet), we also provide an interface (Fig. 1(e)) that allows users (advanced amateur runners or coaches) to manually label biomechanics. The user-customized attribute will then be retrieved from both the sample and exemplar videos for comparison in the same way as those pre-defined attributes. This attribute generalization is facilitated by a design of mappings for biomechanical data that unifies the representations of attributes, their differences and users’ interactions to label the attributes. Specifically, we make use of the semantic model definition of SMPL 3D human mesh model [51]. Users annotate and define measurements on a 3D SMPL body model in T-pose, such that the defined attributes can be retrieved across multiple videos using model correspondence.

We design a user study and expert interviews to evaluate the design components and the overall effectiveness of our system. For the scope of the user study we focus on adults in moderate-speed running (jogging), since this is the most common type and demography for running exercises. The human pose analysis model in VCoach can generalize to other user groups than adult amateur runners, such as children and professional runners, with adapted visualizations of pose differences (e.g., cartoons for children and infographics with detailed figures for professional runners). By replacing the current pre-defined attributes with key attributes of other sports, VCoach can also be generalized to support the posture analysis of new techniques, such as in skating [12] and high jump [38].

2 Related Work

Pose Coaching Systems Previous research work on video-based running pose analysis is limited, partly because in-the-wild running poses contain larger variations in appearance than other sports with more confined locomotion ranges, such as yoga [6] and golf [5]. Running dynamics, such as ground contact time and vertical oscillation, require specific combinations of hardware to capture (e.g., [52]). In the following, we review posture coaching systems in general sports.

According to how the bespoke knowledge of a specific sport is introduced into the system, existing coaching tools span the spectrum from fully-manual to fully-automatic, as illustrated in Fig. 3. The other dimension is whether the poses are captured in 2D (videos) or in 3D (MoCap or Kinect). The fully-manual coaching tools require human coaches to either manually annotate on video playbacks to suggest improvements [37, 48], or analyze data of running gaits captured by MoCap [45]. MotionPro [35], supports manual selection of keypoints on each of the video frames such that some quantities, such as ball trajectory and 2D angles, can be obtained to facilitate analysis. Kinovea [22] and OnForm [37] further simplify the manual tracking by providing basic processing of videos (e.g., automatically track objects and estimate 2D human poses).

![Fig. 2. A toy example showing the faithfulness problem of 3D attributes when projection to 2D. For an 3D angle formed by two 3D vectors (blue and orange), its appearance in 2D is largely different in both vector lengths and angle when observed in different viewpoints.

Fig. 3. The spectrum of level of interference by professionals in existing coaching tools (photo credits from the original sources). The dashed lines indicate range spans on the spectrum.

On the automatic (right) side of the spectrum, a few video-based coaching tools assess the movements based on the reconstructed 2D poses from videos using embedded rules for a specific type of sports, such as skiing (AI Coach) [50] and yoga [6]. Such systems would require extensive domain knowledge to design. To avoid bespoke knowledge, some systems compute suggestions based on the comparisons between novices’ actions with experts’ reference actions. For
example, MotionMA [49] and ReactiveVideo [11] align the experts’ poses captured by Kinect onto the novices’ poses in videos to visualize the difference in postures. AIFit [13] mines and highlights the most significantly different features from the comparisons of reconstructed 3D poses from videos. Even though AIFit is fully automatic, the dominant differences might not reflect informative feedback to the sport.

VCong closes the gap in both dimensions in this spectrum: the input is monocular videos such that it removes the constraint of indoor controlled environments, but it analyzes and visualizes in 3D to ensure spatial awareness. It automatically performs low-level tasks but allows users the controllability to introduce high-level bespoke knowledge to the system.

Video-based Sports Data Analysis A previous work [8] has classified general video-based sports data analysis into four levels: image level, object level, event level, and tactic level. We adopt the same taxonomy as that in [8] and review w.r.t. video-based human pose data in sports. Image-level analysis mainly includes video effects, such as slow-motion playback and displaying frames side-by-side [48]. Image-level analysis does not involve further image understanding from video frames, and thus the image contents would need to be analyzed manually (e.g., by a human coach). Object-level mainly includes obtaining parameters of a single human instance, such as human pose estimation [50] and motion tracking [22, 29]. In sports videos object-level analysis is often more challenging than that in ordinary videos due to motion blurs, large subject displacements and complex sports poses (e.g., high diving). Prior studies addressing these challenges include adopting sports motion priors [17], collecting sports motion datasets [40], and capturing human motions with multi-modal references [13]. Event-level analysis mainly includes recognition tasks from video streams, such as action recognition [40], action quality assessment [27], and key frame detection [44]. Tactic-level is mainly involved in ball games, such as soccer [43], table tennis [8], and basketball [8], by parsing the movements of athletes and objects from videos. VCoach performs object-level analysis, but it focuses on local pose attributes rather than whole-body poses.

User-Customizability in UI The goal of promoting user-customizability is to generalize to new instances other than those embedded in the systems, without requiring end-users’ explicit programming. For example, in gesture recognition, a few systems, such as KinectScript [16] and Visual Gesture Builder [33] allow users to interactively define gestures by recording a few repetitions. MotionMA [49] and YouMove [1] allow users to define movements via Programming by Demonstration (PbD). Besides gestures and movements instances, other finer analysis tasks involve users’ specification of which body part(s) to analyze. A medical research analysis tool, DeepLabCut [32], allows manual labeling of body parts across animal species for training data-driven models. Kinovea [22] and RealitySketch [44] allow users to manually select points to track on top of videos, and customized joint angles can be further computed from the tracked points. While such keypoint definitions apply to a specific video, in this work we develop a systematic set of mappings for users to customize reusable human pose biomechanics across videos.

3 FORMATIVE STUDY

At the beginning of this project we set out to decide the directions and the scope of a sports coaching system suitable for amateurs, which include but are not limited to runners. We conducted a survey on potential target users to understand their usual ways of obtaining feedback on posture correctness in practising sports (Sect. 3.1). We also interviewed three experts on human locomotion to inform our design (Sect. 3.2). The results of this formative study form a set of design requirements for our system (Sect. 3.3).

3.1 Target User Survey

To investigate the demands of potential target users (amateur sports players), we conducted a survey via the Amazon Mechanical Turk (MTurk). We designed a questionnaire with three questions: (1) “What sport(s) do you frequently practise?” (2) “Have you paid attention to the correctness of your body postures while practising the sport(s)?” (3) “If yes, please describe how you get feedback on the correctness of your postures; if not, please explain why not.” We distributed 120 questionnaires in total, and filtered out obvious spam responses according to the quality of the short answers to question (3). Eventually 70 effective answers were collected. Fig. 4 shows the summaries of responses.

Among the responses, jogging/running accounts for the most, followed by football. Other mentioned sports include those involving posture correctness, such as yoga and swimming. 24.3% of the subjects said they only depended on learned instructions of the actions but obtained no feedback; 21.4% of respondents stated that they got feedback from a coach or peers. Other main feedback includes: 5.7% used outcome (e.g., score) as an indicator of posture correctness, 15.7% used feeling (e.g., tense on lower back) as an indicator, and 8.6% adopted extra training on postures. One respondent said he/she video-recorded the actions when practising gymnastics, and two responses explicitly said that they did not get any feedback since no one was watching. Through this survey we learned that the public has the awareness of the importance of maintaining good postures, and there is a need for accessible posture analysis tools. Based on the survey results, we set the focus of our system to jogging, due to its popularity and the requirement on correct postures to avoid injuries, without needing to consider ball/racket trajectories for instrument sports or tactics for team sports.

3.2 Expert Interviews

In order to understand the process and the key factors of human movement analysis, we conducted semi-structured interviews with three experts, two were medical doctors in Sports Medicine working in a hospital (E1, E2), and the other one (E3) was a researcher in Sports Science in a startup company studying performance analysis in sport. During the interviews we first invited the participants to describe a representative case in which human movement analysis is involved in their daily practice. During the description, they were asked to identify what is the routine they analyze human movements, what are the key factors they focus on, and what is the decision process based on their observations. Then we raised open questions such as difficulties in human movement analysis, and the role of video-based analysis in practice.

All of the three experts mentioned that human movement analysis is based on gold standards, i.e., comparisons with the normal values in rehabilitation exercises or with top athletes’ postures and performances in sports. Even for a full-body movement only a few key factors are concerned in evaluation (deterministic models [17]). For example, E1 described a case of imbalance testing, where the key factors were movement accuracy and time required for completion. E3 emphasized the advantage of externally-focused training over internally-focused training [53]. He pointed out that even though real-time feedback provides direct guidance, it would distract a subject during the action by interfering the subject’s intention of movements. He also mentioned that since a coach’s attention is limited, he/she often can only focus on a specific body part during instruction, and that it would be ideal to analyze other parts during playback.
Since our system is focused on running, throughout the project we closely worked with E3 and another expert (E4), a third-year postgraduate student in Sports Science, who was involved after this formative study. We initiated discussions with them as needed via remote chats.

3.3 Design Requirements

From the expert interviews on the human movement analysis, as well as the limitations of existing systems, we identify the following design requirements:

R1 - The tool should be accessible to users without an expert.

The potential users of our system might have no domain knowledge to determine the posture correctness directly from their videos. This can be mitigated by comparing their videos with another video involving standard running poses from a professional runner and learning from the differences. Our system should not only include key factors to running but should also allow users to easily introduce other key factor(s) in case needs arise, instead of embedding redundant bespoke knowledge of running in the system. Our system should be as easy to use as possible for novice users.

R2 - The comparison should adapt to variations.

The videos input by users may contain large variations on the running poses, due to viewpoints and subjects’ physical characteristics. The comparison should be able to factor out these differences and focus on only factors that indicate running posture correctness.

R3 - The visualization should highlight the part-based differences.

As pointed out by E3, the attention of both coaches and athletes is limited, they are often advised to correct one part at a time. Thus instead of showing all the mistakes at the same time, our system should show the differences in each body part separately. E3 also mentioned that for both coaches and athletes the quantitative figures do not make sense; they desire a direct corrective suggestion. Thus instead of presenting analysis results as infographics, we need to design an intuitive way to demonstrate the differences.

R4 - The system should enable user interactivity.

As suggested by E4 in a later discussion, when a coach corrects an action, he/she usually first points out the mistakes, and then shows the correct action. Our system should also follow this routine. Following the design requirement R1, since there is no remote coach explaining the results, our system should allow users to explore the feedback to make the most sense out of it.

4 System Overview

We design our system VCoach based on the aforementioned requirements. Since we target novice users, the overall system workflow follows the “overview first, details-on-demand” principle [41]. Users input videos and preview suggestions through the user interface (Fig. 1). The input to our system contains two videos (Fig. 1(a)): a sample running video to be analyzed, and an exemplar running video for comparison (R1). Upon loading the two videos, our system automatically processes the videos to reconstruct 3D human poses, normalizes the motions (R2), and segments the videos into running cycles. Our system then performs the pose analysis by aligning the sample and exemplar running pose sequences based on 3D pose similarity, and retrieves the pre-defined key attributes to conduct comparisons (R1). The suggestions for correction are generated based on the part-based differences from the comparison (R3), and directly reflect on a timeline tailored for running pose sequences (Fig. 1(c)). Those attributes that require improvement are represented with glyphs. By clicking on each glyph on the timeline (R4), a detailed instruction for improving the corresponding attribute is shown as a short 3D animation of a body part on a human model in the suggestion preview window (Fig. 1(d)). Users can rotate the body model to navigate through viewpoints for better perception (R4).

For other pose attributes that are not embedded in our system as pre-defined attributes, the users can interactively label (R4) on a 3D body model via the query editor (Fig. 1(e)). The labeled attributes will then be retrieved and analyzed from the videos in the same way as the pre-defined attributes. Our system contains five modules, as shown in Fig. 5.

5 Visual Design

In this section we first describe the formulation of the data (attributes of running poses) we study to design our system. Then we propose three mappings based on the data, i.e., the representation of data, the visualization of their differences, and the user operations to interactively define the attribute of each type.

5.1 Data Attributes

The data attributes in our system include both pre-defined attributes that are commonly used for evaluating running poses, and user-defined attributes for their customized analysis. To determine the common data attributes for running pose correction, we collected a corpus of running pose tutorials by searching with key words “running pose tutorials”, “running pose corrections”, “running techniques”, “running form”, etc., from Google and YouTube. The current corpus contains 55 items (including 37 videos and 18 articles). The data attributes are summarized from the corpus into four types, as shown in Fig. 6.

We conducted another interview with E4 to verify the coverage of these attributes in running pose evaluation in practice. The fourth type “categorical data” is different from the previous three in that they are not computed from comparison with exemplar poses, but computed directly based on the other three classes (i.e., first compute a value and then discretize it into a category by a certain threshold). Thus we focus on the design for the first three types, but support the visualization of the categorical data for commonly evaluated attributes in running.

Fig. 5. The architecture of VCoach, which comprises five main modules.

Fig. 6. The data attributes for evaluating running poses summarized from our collected corpus. In the brackets are the numbers of their occurrences in the corpus.
5.2 Representing attributes

In this section we summarize the visual encoding of the positional, angular, and temporal attributes.

Positional attributes (Fig. 7(a)) are defined as the relative distance between two points (classified as type P1), or the position of a point from a specific axis (P2). For example, the trajectory of the wrist is its relative distance to the body center (P1). Another example is the knee lift, which is a vertical distance from the knee joint to the body center (P2).

Angular attributes (Fig. 7(b)) are defined as either the angle formed by three endpoints (classified as type A1), or the orientation of a vector formed by two joints with respect to an axis (A2). For example, the elbow angle (A1) is an angle formed by the shoulder, the elbow and the wrist joint. The leaning of the upper body (A2) is the orientation of the vector pointing from the root joint to the neck joint w.r.t. the z-axis.

Temporal attributes are defined as either a single moment (T1) or a time range within a running cycle (T2). We use a temporal axis to show the temporal context. The temporal axis (Fig. 7(c)) is a fixed full running cycle, with the three dots from left to right respectively corresponding to the states of right foot landing (RL), left foot landing (LL), and right foot landing for the next cycle. The positioning of the human center on the temporal axis reflects the state of the current pose within the running cycle.

5.3 Visualizing attribute differences

This section introduces the representation of the differences in data attributes. Such differences are mainly used for presenting feedback, i.e., from an incorrect configuration to a correct one. We define a set of visuals for attribute differences (Fig. 7(d)), which are unified with the attribute representation. Positional difference is shown by two points and an arrow pointing from the wrong position to the correct position. Angular difference is shown by two vectors forming a wedge to show an angular difference. Temporal difference is represented by a red marker segment on the temporal axis showing a temporal offset. For example, the red segment along the forward temporal axis direction indicates the current event should appear later.

5.4 User operations

In this section we introduce the user operations (Fig. 7(e)) for defining their own data attributes under the three data attribute classes. Specifically, the query editor in our user interface (Fig. 1(e)) contains a 3D viewer presenting the 3D human body model in T-pose, radio buttons for specifying properties and two draggable cursors (red lines) on top of a running cycle diagram for specifying timings. A user may either refer to the mesh or skeleton of the body model and directly mouse-click on the body model to select joints; our system will snap the mouse click to the nearest joint.

A user first selects the attribute type by selecting either the angle button or distance button for angular and positional attributes, respectively, or directly dragging the temporal cursors for a temporal attribute. To edit a positional attribute, a user first specifies the joint to track, and then specifies the base point (P1). When the user further selects an axis, only the component of the selected dimension will be recorded (P2). To edit an angular attribute, a user either selects three endpoints in order on the body model (A1), or two points and one axis (A2). To edit a temporal attribute, the user either moves one cursor to specify a moment (T1), or both cursors to specify a time range (T2). Our system will record a phase or a phase range accordingly. When the positional and angular attributes are associated with an event, the user also moves the temporal cursor to specify the timing. Please refer to the demo video for the authoring process of “left foot landing position” example.

5.5 Design of Attributes Overview

In this section we discuss the design of the overview for the problems reflected from the comparison. The overview should show which attributes appear in question in the sample video and their timings. We
Thus propose to use glyphs for representing attributes and a timeline tailored for running to organize them temporally.

**Glyphs** We designed two types of glyphs for the four classes of the attributes, namely suggestion glyphs and profile glyphs. Suggestion glyphs are icons for each of the three classes of attributes in Fig. 7 i.e., positional, angular and temporal attributes in the collected corpus, whose values are continuous variables and are compared with those in the exemplars. As shown in Fig. 8(a-c), the suggestion glyphs are designed based on the idea of traffic signs that augment markers to symbols, such that users do not need to memorize the encoding, but can easily get familiar with the meaning of the icons and can interpret the meaning by intuition.

The profile glyphs are used to represent categorical attributes which do not need comparison with the exemplar. We adopt the idea from the dance notations [34] to discretize complex human movements into reference planes (sagittal, frontal and horizontal). As shown in Fig. 8(d), we use three transverse planes that capture the joints with a large degree of freedom, i.e., foot, knee, and shoulder. Then the motions of these joints in relation to the body center are reflected by their projections into the three transverse planes. In this figure, the left shows a mid-foot strike while the right shows a fore-foot strike at landing. We assign another glyph (e) to represent a customized attribute on the timeline.

The profile glyphs are used to represent categorical attributes which do not need comparison with the exemplar. We adopt the idea from the dance notations [34] to discretize complex human movements into reference planes (sagittal, frontal and horizontal). As shown in Fig. 8(d), we use three transverse planes that capture the joints with a large degree of freedom, i.e., foot, knee, and shoulder. Then the motions of these joints in relation to the body center are reflected by their projections into the three transverse planes. In this figure, the left shows a mid-foot strike while the right shows a fore-foot strike at landing. We assign another glyph (e) to represent a customized attribute on the timeline.

**Timeline** A characteristic of a running pose attribute sequence is temporally periodical, and each period can be divided into a right-phase and a left-phase. Based on this characteristic, we propose to design a timeline that transforms the temporal space into a running event space. As shown in Fig. 1(c), the horizontal axis is a complete running cycle, and the vertical axes correspond to the attributes of the left side of the body, right side of the body, and middle, respectively. All the data attributes are summarized among cycles to be shown on the timeline. Our system will automatically select significant errors, with the sizes of the glyphs proportional to the significance of the errors of a particular type.

We have conducted a pilot study to verify the above design against their alternatives. For glyph design, the alternatives include a set of simplified icons highlighting the body parts in question, and color and shape encoding. For timeline design, the alternatives are an ordinary linear timeline of a video that is not segmented into running cycles, and a spiral timeline displaying all running cycles without summarizations. We invited two users, both of them are novices to running, and one of them with design background. We introduced the overall function of our system along with the two sets of designs, and then let them vote on which representation they prefer. Both of them chose the semantic glyph and the aggregated timeline, because they thought the semantic icons are intuitive and can be easily remembered. As novice users they do not desire all the occurrences of the problems, but rather what kinds of problems appear in their running; thus the aggregated timeline is more preferable.

**Fig. 9.** Alternative designs for: (a) icons and (b) timeline.

### 6 Data Analysis Model

In this section we introduce the methods of the backend modules in **VCoach** (Fig. 5): video processing, pose analysis, and feedback.

#### 6.1 Video Processing

**3D Pose Reconstruction and Normalization** When the sample and the exemplar videos are loaded into the system, the pose at each frame is retargeted onto the SMPL models, denoted as \( M_s \) for the sample video and \( M_e \) for the exemplar video. The retargeting (reconstruction) is implemented with TCMR [9], which is a monocular pose reconstruction method achieving state-of-the-art accuracy on challenging outdoor video datasets. \( M_s \) and \( M_e \) are then rotated to a unified global orientation to facilitate comparison (Fig. 1(a)). The video frames are cropped to maximize their preview in windows.

**Temporal Segmentation** The running pose sequences in both the sample and exemplar videos are segmented by the key frames of foot landing and foot extension. Since the action of running is periodical, we adopt the phase variable of human locomotion, as in [20]. A full running cycle thus contains four key phases, in “right foot landing” (phase = 0), “right foot extension” (phase = 0.25), “left foot landing” (phase = 0.5), and “left foot extension” (phase = 0.75) order. These four key phases are detected from the local extrema of the foot trajectories.

#### 6.2 Pose Analysis

**Sequence Alignment** Given the detected key phases, the running pose sequences in the sample and exemplar videos are first temporally aligned at key phases, and then aligned at a finer level between each two key phases using the dynamic time warping technique [4]. We use joint rotations to measure human pose similarity [30].

**Attributes Retrieval** Each data attribute is represented as a meta data tuple: \( \text{name.type}, J_A, J_B, \text{axis}, \text{side}, \text{phase}) \), where \( J_A, J_B \) are the joint IDs of the attribute endpoints in the body model (as shown in Fig. 7); \( \text{side} \) is one of the “left”, “neutral” and “right”; \( \text{axis} \) and \( \text{phase} \) are the related axis and timing of the attribute; they are left empty if not
The main purpose of the user study is to evaluate the improvement in animated suggestions w.r.t. specific 3D local pose attributes. We develop a viewpoint navigation and summarization of feedback) in assisting novices’ perceptions of running pose improvements.

Comparison Since different attributes have different scales and units, we normalize the attribute values to the range [0, 1]. Then the differences in the attribute values are computed as the relative errors between the attributes from the sample video and those from the exemplar video. We set a threshold of 25% to select significantly different attributes and scale the sizes of the suggestion icons according to the relative errors.

6.3 Feedback

Animation-based Demonstration The corrective suggestion from pose comparison is conveyed by animating a 3D human model. To make the demo easily understandable, the animation follows the design guideline as data-GIF [42]. The animation contains two key frames corresponding to the wrong pose and the same pose with a specific body part in the position as the exemplar pose, respectively. Specifically, we use the joint rotations to drive the model: for angular attributes, the intermediate frames are interpolated with the joint rotations of \( J_a \); while for positional attributes, the animation is interpolated with the joint rotations of the parent joint of \( J_a \) along the kinematics tree. The 3D animations are augmented with visual markers to highlight differences, as in Fig. 7(b).

Suggestive Viewpoint Since the animation of corrective suggestions is in 3D, we would like to demonstrate it at the most informative viewpoint.

While there are prior studies on the automatic selection of viewpoints for previewing a 3D mesh, the definition and criteria of the optimal viewpoints are often dependent on the purpose, such as to demonstrate region visual saliency [25], to set man-made models in upright orientation [14], and to incorporate modelers’ creation processes [5]. Previous studies on optimal viewpoints for human poses mainly include reducing prediction uncertainty in estimating 3D pose [21] and metrics defined over body part visibility [25]. In \( VCoach \), since we would like to provide suggestions w.r.t. specific 3D local pose attributes, we develop a set of schemes to suggest viewpoints according to the geometry of the attributes.

The main idea is to minimize the ambiguity in the attributes due to camera projection, while preserving the human model as the spatial context. Based on this goal, we make use of the normal vector formed by the 3D attributes to decide the orientation of the viewpoint (see Fig. 10). We further use the side of the body to determine whether to revert a normal to its opposite direction. For example, to present an attribute on the right side of the body, the camera should also be placed to the right facing the body model. The up direction of the viewpoint is along the average of the two vectors. We also determine whether to revert the up direction according to whether it keeps the human model heading upwards. Even though we present the 3D animation in the suggested viewpoint, users can still manually change the viewpoint to explore the corrective suggestion.

7 Results

In this section, we show the results of a user study evaluating the visualizations of posture correction feedback in \( VCoach \) and the baseline methods (Sect. 7.1) for novices, and expert interviews (Sect. 7.2) to evaluate the overall effectiveness of the system in pose correction.

7.1 User Evaluations

The main purpose of the user study is to evaluate the improvement of \( VCoach \) in promoting novices’ perception of running pose differences over existing methods (see Baselines). It also evaluates the effectiveness of other components (e.g., viewpoint navigation and summarization of feedback) in assisting novices’ perceptions of running pose improvements.

Fig. 10. Deciding viewpoints by the geometry of the 3D attributes. (a) For angular attributes the viewpoint is along the normal of the plane where the angle lies in, pointing outwards the body model; (b) for positional attributes the viewpoint is along the normal of the plane formed by the wrong position, the correct position, and the body center; (c) the temporal axis is within the Sagittal plane of human, and the viewpoint is perpendicular to this plane.

Fig. 11. The UI for the baseline methods: (a) juxtaposition: compare two synchronized frames side-by-side; (b) superposition: overlay the transformed pose in the exemplar frame onto the sample frame.
explore the corrective feedback from videos using either VCoach or the baseline methods in a think-aloud manner, and complete a questionnaire afterwards. The user study contained three sessions: two sessions using our system with and without the suggestive viewpoints, and one session using the baseline methods. The order of the three sessions was counter-balanced, and the order of the nine videos was randomized among the three sessions (three videos for each session). During training, we first gave a detailed tutorial on the operations of VCoach as well as the baseline system. The participants then tried freely to get familiar with both systems.

In the session using VCoach without suggestive viewpoints (denoted as “VCoach-w/o”), we disabled the suggestive viewpoint function, and the participants would need to manually navigate the viewpoints to preview the 3D animations. The system recorded the participants’ navigation activities in the suggestion preview window, parameterized by viewpoint azimuth and elevation, and the duration of each viewpoint. In another session using VCoach (denoted as “VCoach”), the suggestive viewpoint function was enabled; the participants could also manually navigate, and their navigation activities were also recorded. In the session using the baseline methods (denoted as “Baseline”), the participants explored the corrective feedback by comparing running poses in videos in either juxtaposition or superposition visualization.

After the sessions, the participants completed a designed questionnaire in a 7-point Likert Scale (1 is Strongly Disagree and 7 is Strongly Agree), and a standard System Usability Scale (SUS) [4]. The user study with each participant took about 90 minutes.

| Q1 | The feedback of posture correction is easy to access. |
| Q2 | The demonstrations of pose differences are easy to understand. |
| Q3 | The visual designs are intuitive. |
| Q4 | The feedback reflects the problems in sample videos. |
| Q5 | The feedback is helpful in improving running postures. |
| Q6 | Demonstrations with animation, normalized poses, summary of mistakes, suggested viewpoints are helpful for understanding suggestions. |
| Q7 | I’m more satisfied with VCoach than only browsing videos and overlaid poses. |

Table 1. The questionnaire used for the user study: Q1–Q5 evaluate the effectiveness of VCoach in providing corrective feedback on running poses. Q6–Q10 evaluate the advantages of design components in VCoach over the baseline methods.

![Fig. 12. Summaries of the participants’ responses to the questionnaire in Table 1. The black vertical bars represent the median of the ratings for each question.](image)

**Effectiveness of Pose Difference Visualization** We first investigate the effectiveness of VCoach in presenting feedback compared with the baseline system. Q10 explicitly asked the comparison between VCoach and the baseline methods, where 10 out of 12 participants strongly agreed that VCoach was more effective in conveying feedback than the baselines. We recorded the time required to explore the running pose problem(s) in each video, as shown in Fig. 13(a). Paired t-tests on exploration time required for each video between sessions “VCoach” and “Baseline” showed that using VCoach with the suggestive viewpoint significantly requires less time to obtain the desired feedback ($p = 0.019$). However, there is no significance on exploration time between sessions “VCoach-w/o” and “Baseline” ($p = 0.519$).

We evaluated the accuracy via the successful rate of the participants’ discovered mistakes matched the ground-truth mistakes as commented by the coaches in videos. In sessions “VCoach-w/o” and “VCoach” the successful rate was 100%. In other words, all the participants could figure out the problem(s) in the running poses with the visualization provided by VCoach. In contrast, the successful rate was 77.8% in session “Baseline”. From the participants’ think-aloud in session “Baseline”, they often referred to the superposition visualization more than the juxtaposition visualization, especially when the subjects in the sample and exemplar videos are running in different directions. For superposition in the baseline system, a6 and a8 said that they would refer to the lower limbs more often than upper limbs, since upper limbs were often occluded and misaligned due to differences in limb lengths.

![Fig. 13. Statistics on time in the user study. (unit: s). (a) Average exploration time for using the baseline methods, VCoach without suggestive viewpoint, and VCoach with suggestive viewpoint. (b) Average editing time for the three selected attributes using the query editor.](image)

**Effectiveness of System Components** We then investigate the influence of specific design components on users’ perception of feedback on running pose correction. Q6 asked the participants to rate the key component in VCoach, which visualizes pose differences via animations of local body parts on a human model. 8 out of 12 participants strongly agreed that such visualization was helpful for understanding, and the other four chose agreed. The component that received the most disagreement is the preview of normalized poses from the sample and exemplar videos shown in juxtaposition (Fig. 1(a) middle). Since their orientations are often different from those in the original videos, the participants stated that referring to them increased the cognitive load by having to imagine the transformation to understand. Thus even though normalized poses are crucial to computing pose differences, they do not necessarily contribute to users’ visual comparison. During the participants’ think-aloud in sessions “VCoach-w/o” and “VCoach”, they often directly moved on to check the glyphs on the timeline after loading both videos. After watching the animation, they sometimes checked the sample video frame to verify the problem. At first they sometimes also referred to the exemplar frame to verify the animation, but many of them skipped the exemplar frame later because they found the corrective feedback illustrated by the animation was trust-worthy.

We also evaluated the usefulness of the design component of suggestive viewpoint. We would like to figure out the following two questions: (1) do users find previewing the animations of pose correction under a certain viewpoint yields better perception? (2) If yes, do our suggestive viewpoints match the preferred viewpoints selected by users? We thus analyze the usage of viewpoint selection during the user study. In session “VCoach-w/o”, the average number of times the participants manually changed the viewpoint was 7.36 times per video, compared with 2.05 times per video in session “VCoach”. A paired t-test on the numbers of manual navigation between sessions “VCoach-w/o” and “VCoach” shows that enabling the suggestive viewpoint function...
significantly reduces users' manual navigation ($p = 0.00059$). To answer question (2), we further analyze the relevance of the participants' manually-selected viewpoints with the suggested viewpoints computed by our system in session “VCoach-w/o”. We analyzed previewing viewpoints that lasted more than one second and considered those with a duration less than one second as the navigation process. The average errors of azimuth and elevation relative to 360° were 3.19% and 4.52%, respectively, indicating a good match between our suggestive viewpoints and preferred viewpoints by the participants.

In the rating of the usefulness of suggestive viewpoint, seven participants chose “strongly agree”, and four of them explicitly stated during exploration that this function was very convenient. a2 in session S1 asked whether the suggestive viewpoint function could be enabled, because she found this function especially useful when she was comparing the magnitudes of corrections on foot landing position. a4 found the suggestive viewpoint more useful in observing upper limbs because they often suffer from heavier occlusions by the body torso than lower limbs. Interestingly, a12 rated “Neutral” in Q9. He explained that since he studied exoskeleton robotics, he was more used to imagining the attributes using the sagittal, coronal and transverse planes as reference, rather than using the human body as a spatial context. Since VCoach targets at novice users without human movement analysis background, and most participants found the suggestive viewpoint function convenient, it can serve as a helpful option in VCoach.

System Usability In the training session, all the participants could get familiar with VCoach within 5 minutes by completing a pipeline of operations, including loading videos, previewing frames and poses, and navigating on the timeline to preview animations of suggestions. The SUS score for all the ten questions in the SUS questionnaire was 83.125 on average (SD: 10.36), out of a scale of 100, indicating the good usability of VCoach.

In post-study interviews with the participants, they commented favorably towards VCoach. For example, a3: “Besides clarity, the summarization in VCoach helps me form a better impression of frequent mistakes. With VCoach I don’t even have to browse the entire video, but only need to refer to the frames the system has highlighted for me.” The participants also commented on the potential generalization of VCoach in other scenarios. Specifically, a11: “This tool is solving a very practical problem. I can see how it is useful in running and can imagine it generalizes to many other sports.” a12 (from exoskeleton robotics background): “…current rehabilitation training often relies on wearable sensors to detect patients’ biomechanics, such as joint angular velocities and accelerations. Such a video-based tool is promising in providing a non-invasive means to analyze patients’ movements.”

Evaluation of Query Editor From the user study we also evaluate the easiness of use of the query editor, specifically, how efficiently and accurately users can edit a pose data attribute. There is no baseline method for this task. We chose three frequently used data attributes from each of the classes in the pre-defined attributes, and asked the participants to edit the attributes using the query editor in our interface. The three attributes were: “foot landing position” (P2), “elbow angle” (A1) and “foot contact time” (T2). They covered all the operations on the query editor. The participants were given sample running video clips as references. As shown in Fig. 13(b), the average editing time for the three attributes were 95.36s (SD = 37.71), 39.91s (SD = 10.11) and 38.64s (SD = 14.03). On average the editing of the foot landing position took the longest time, since it required the most operations covering all the components on the query editor. The successful rates that the participants can implement the same attribute as our pre-defined was 83.3%, 100%, and 91.7%, respectively. In the failure cases, a3 failed the temporal attribute, because he misunderstood the question and labeled the time between two consecutive foot landings instead. a4 and a10 both correctly annotated the positional attribute on the human model, but forgot to associate with the timing for foot landing by dragging the timeline cursor. Through this experiment we verified that novice users could easily understand and implement the representative attributes with minimal training. Even though for most amateur runners the pre-defined attributes would suffice, they can annotate their interested attributes via the query editor with reasonable efforts.

7.2 Expert Interviews We conducted expert interviews to evaluate the overall usefulness of our system in helping amateur runners correct running poses. Two experts with running backgrounds were invited: one was a licensed running coach (E5); the other was a professional marathon runner (E6). The two interview sessions were conducted separately, and each session lasted 50 minutes. During the interviews we provided a detailed introduction of functions in VCoach with three demonstrations of usage scenarios, and then invited them to try the system freely.

Both experts strongly agreed that VCoach would benefit a lot of runners. E5: “Not only beginners, but experienced runners are also often bothered by the problems of running pose correctness. I can expect this tool serves a lot of runners.” They also appreciated that the design rationale of VCoach is very reasonable for practical usage. E5 said that coaching is a highly personalized process; and thus there is no absolute “correct” running pose regulated by numbers, such as the legal range of elbow angle in degree. A significant advantage of the design of VCoach is that it does not directly classify a runner as right or wrong, but retains the flexibility to compare with various running poses to show the differences. E5 thus finds VCoach especially useful for novices to iteratively adjust to different exemplars to find their most suitable poses. E6 commented that the design of VCoach is similar to the idea of the “champion model” for elite athletes, such as Su Bingtian, who was trained by shortening the gaps (on both poses and capabilities) with elite exemplars. This comment is consistent with E3’s advice in the formative study.

We also invited experts to comment on the positioning of VCoach in training in real life. E5: “It is suitable for the majority of ordinary runners. But for severely over-weight people, asking them to resemble the running of ordinary people might cause injury instead of reducing it; they should seek for professional advice instead.” E6 suggested that if the athletes’ parameters (mainly including height, leg lengths and training years) in the videos are accessible, it would be helpful to also suggest exemplars to users according to the similarity in these parameters, since runners with similar body configurations are more likely to have similar suitable running poses.

8 Conclusion and Future Work We have presented a novel system, VCoach, for assisting amateur runners in improving their running poses. We designed the system based on the design requirements formed from the literature research and expert interviews. VCoach embeds common running pose attributes based on a collected corpus, and also provides an interface for users to customize attributes. VCoach analyzes the poses from a sample video and an exemplar video in 3D, and visualizes the pose differences via 3D animations on a human body model. Our user study showed that demonstrating pose corrective feedback via 3D animations is more effective than displaying frames side-by-side or overlaying the correct poses onto the sample frames.

There are several limitations and possible future work directions for VCoach. In the current setting the running pose attributes are analyzed and visualized independently. But there are certain correlations among the attributes, e.g., a higher knee lift might yield a larger stride. A potential improvement is to incorporate human body harmonics [23,24] to further summarize the problematic attributes. Besides, in our user study we mainly evaluated the effectiveness of the visualization in VCoach in providing intuitive pose correction feedback. It would be meaningful to conduct a long-term user study with participants from running backgrounds to further evaluate the effectiveness of VCoach in promoting running forms in practice. Finally, currently VCoach focuses on the kinematics measurements (e.g., angles and positions). However, more professional analysis [31] would require kinetics measurements, such as ground reaction force (braking force) [33] and muscle elastic energy [39]. Since the measure of kinetics parameters is currently limited to biomechanics laboratories, developing methods that recover the kinetics from videos would increase accessibility to many fields, including but not limited to sports posture analysis.
[47] H. Techiques. Hudl: Performance analysis tools for sports teams and athletes at every level, 2007.

[48] TechSmith. Coach’s eye. https://www.coachseye.com 2011.

[49] E. Velloso, A. Bulling, and H. Gellersen. Motionma: motion modelling and analysis by demonstration. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 1309–1318, 2013.

[50] J. Wang, K. Qiu, H. Peng, J. Fu, and J. Zhu. Ai coach: Deep human pose estimation and analysis for personalized athletic training assistance. In Proceedings of the 27th ACM International Conference on Multimedia, pp. 374–382, 2019.

[51] R. X. Wei, Z. Y. Chan, J. H. Zhang, G. L. Shum, C.-Y. Chen, and R. T. Cheung. Difference in the running biomechanics between preschoolers and adults. Brazilian Journal of Physical Therapy, 25(2):162–167, 2021.

[52] P. W. Woundefinedniak, M. Zbytniewska, F. Kiss, and J. Niess. Making Sense of Complex Running Metrics Using a Modified Running Shoe. 2021.

[53] G. Wulf, C. Shea, and R. Lewthwaite. Motor skill learning and performance: a review of influential factors. Medical education, 44(1):75–84, 2010.

[54] G. Yan and M. Woźniak. Accurate key frame extraction algorithm of video action for aerobics online teaching. Mobile Networks and Applications, pp. 1–10, 2022.

[55] D. Zecha, C. Eggert, M. Einfalt, S. Brehm, and R. Lienhart. A convolutional sequence to sequence model for multimodal dynamics prediction in ski jumps. In Proceedings of the 1st International Workshop on Multimedia Content Analysis in Sports, pp. 11–19, 2018.