Qualitative Activity Recognition of Weight Lifting Exercises

Eduardo Velloso
Lancaster University
Lancaster, UK
e.velloso@lancaster.ac.uk

Andreas Bulling
Max Planck Institute for Informatics
Saarbrücken, Germany
andreas.bulling@acm.org

Hans Gellersen
Lancaster University
Lancaster, UK
hwg@comp.lancs.ac.uk

Wallace Ugulino
Pontifical Catholic University of Rio de Janeiro
Rio de Janeiro, Brazil
wugulino@inf.puc-rio.br

Hugo Fuks
Pontifical Catholic University of Rio de Janeiro
Rio de Janeiro, Brazil
hugo@inf.puc-rio.br

ABSTRACT

Research on activity recognition has traditionally focused on discriminating between different activities, i.e. to predict "which" activity was performed at a specific point in time. The quality of executing an activity, the "how (well)", has only received little attention so far, even though it potentially provides useful information for a large variety of applications. In this work we define quality of execution and investigate three aspects that pertain to qualitative activity recognition: specifying correct execution, detecting execution mistakes, providing feedback on the to the user. We illustrate our approach on the example problem of qualitatively assessing and providing feedback on weight lifting exercises. In two user studies we try out a sensor- and a model-based approach to qualitative activity recognition. Our results underline the potential of model-based assessment and the positive impact of real-time user feedback on the quality of execution.

Categories and Subject Descriptors

H.5.2 [**Information interfaces and presentation**]: Miscellaneous.; I.5.2 [**Pattern Recognition: Design Methodology**]: Feature evaluation and selection

General Terms

Algorithms, Human Factors

Keywords

Qualitative Activity Recognition, Weight Lifting, Real-Time User Feedback

1. INTRODUCTION

It is well-agreed among physicians that physical activity leads to a better and longer life. For example, a recent consensus statement from the British Association of Sport and Exercise Sciences showed that physical activity can reduce the risk of coronary heart disease, obesity, type 2 diabetes and other chronic diseases [24]. Moreover, a recent study estimated that at least 16% of all deaths could be avoided by improving people's cardio-respiratory fitness [5]. An effective way of improving cardio-respiratory fitness is to regularly perform muscle strengthening exercises. Such exercises are recommended even for healthy adults as they were shown to lower blood pressure, improve glucose metabolism, and reduce cardiovascular disease risk [24].

A key requirement for effective training to have a positive impact on cardio-respiratory fitness is a proper technique. Incorrect technique has been identified as the main cause of training injuries [13]. Moreover, free weights exercises account for most of the weight training-related injuries (90.4%) in the U.S. [19]. The same study states that people using free weights are also more susceptible to fractures and dislocations than people using machines. The predominant approach to prevent from injuries and provide athletes with feedback on their technique is personal coaching by a professional trainer. While highly effective, the presence of a trainer may not always be possible due to cost and availability. Personal supervision also does not scale well with the number of athletes, particularly among non-professionals.

A particularly promising approach to assessing exercises and to providing feedback on the quality of execution is to use ambient or on-body sensors. In sports science, a standard approach employed by trainers is to film the athlete using a camera and to use a video digitising system to perform offline frame-by-frame annotation of the data. Alternatively, athletes can use marker-based tracking systems that automatically generate a digital skeleton. In activity recognition using on-body sensing, a large body of work has investigated automatic techniques to discriminate *which* activity was performed. So far, only little work has focused on the problem of quantifying *how (well)* an activity was performed. We refer to the latter as "qualitative activity recognition".

The aim of this work is to investigate the feasibility of automatically assessing the quality of execution of weight lifting exercises and the impact of providing real-time feedback to the athlete - so-called *qualitative activity recognition*. We

AH'13, March 07 - 08 2013, 2013, Stuttgart, Germany.

focus on three aspects that we believe are key components of any qualitative activity recognition system, namely the problem of specifying correct execution, the automatic and robust detection of execution mistakes, and how to provide feedback on the quality of execution to the user. More specifically, we explore two approaches for detecting mistakes in an automated fashion. The first is to use machine learning and pattern recognition techniques to detect mistakes. The second approach, and the one proposed in this paper, is to use a model-based approach and to compare motion traces recorded using ambient sensors to a formal specification of what constitutes correct execution.

The specific contributions of the work are: (1) a formalisation of the term "quality" in the context of activity recognition; (2) the design and implementation of a novel framework for the development of qualitative activity recognition systems; and (3) the evaluation of a system developed with the framework in a user study on the example problem of assessing the quality of execution of weight lifting exercises.

2. RELATED WORK

2.1 Recognition of Sports Activities

A large number of researchers have investigated means to provide computational support for sports activities. For example, Michahelles *et al.* investigated skiing and used an accelerometer to measure motion, force-sensing resistors to measure forces on the skier's feet and a gyroscope to measure rotation [21]. Ermes *et al.* aimed to recognize several sports activities based on accelerometer and GPS data [12]. In the weight lifting domain, Chang *et al.* used sensors in the athlete's gloves and waist to classify different exercises and count training repetitions [9]. More recently, the Microsoft Kinect sensor has been used in research and uses a depth camera to extract a skeleton [11], which shows great potential for tracking sports activities unobtrusively.

2.2 Qualitative Assessment

While several works explored how to recognize activities only few addressed the problem of analysing their quality. There has been work on using cameras for tracking spine and shoulders contours, in order to improve the safety and effectiveness of exercises for elder people [16]. Moeller *et al.* used the sensors in a smartphone to monitor the quality of exercises performed on a balance board and provided appropriate feedback according to its analysis [22]. Similarly, Wii Fit is a video game by Nintendo that uses a special balance board that measures the user's weight and center of balance to analyse yoga, strength, aerobics and balance exercises, providing feedback on the screen. With the objective of assessing the quality of activities Hammerla *et al.* used Principal Component Analysis to assess the efficiency of motion, but focused more on the algorithms rather than on the feedback [15]. *Strohrmann et al.* used inertial measurement units installed on the users' foot and shin to analyse their running technique, but didn't provide feedback either [30].

2.3 Model-based Activity Recognition

Because sports exercises are often composed of well-defined movements, it is worth analysing approaches that leverage the capabilities of a model to analyse activities. For example, Zinnen *et al.* compare sensor-oriented approaches to model-based approaches in activity recognition [31]. They proposed to extract a skeleton from accelerometer data and demonstrated that a model-based approach can increase the robustness of recognition results. In a related work, the same authors proposed a model-based approach using high-level primitives derived from a 3D human model [32]. They broke the continuous data stream into short segments of interest in order to discover more distinctive features for Activity Recognition. Reiss *et al.* used a biomechanical model to estimate upper-body pose and recognize everyday and fitness activities[26]. Finally, Beetz *et al.* used a model-based system to analyse football matches in which players were tracked by a receiver that triangulated microwave senders on their shin guards and on ball [4].

2.4 User Feedback

Some works that include feedback to the athlete include displaying performance statistics on a screen for rowing, table tennis and biathlon training [1]. Iskandar *et al.* even proposed a framework for designing feedback systems for athletes [18]. Hey *et al.* used an enhanced table tennis practice table to visualize past impact locations by tracking the ball using a video camera and a vibration detector [17]. A few works have explored how to provide feedback on swimming technique using a GUI [23] and a multimodal approach [2]. Several works aimed to track exercises to provide feedback and thus increase motivation. Examples include the commercial Nike + iPod that combines data gathered from sensors in the user's shoes with music, MPTrain that builds a playlist by using the mapping between musical features, the user's current exercise level and the physiological response [25], and MOPET that uses GPS, acceleration and heart rate data to increase motivation and provide advice to the user through a 3D avatar on a mobile device [8].

There has also been work on using sensors to provide physical activity energy expenditure, since the amount of calories burnt in an exercise is a very important metric for performance evaluation. Approaches in this sense include SensVest, a wearable device to record physiological data from children playing sports [20] and using artificial neural networks to estimate energy expenditure [29].

3. QUALITY IN ACTIVITY RECOGNITION

In order to discuss qualitative activity recognition we first need to define what we mean by the "quality of an activity". Although some works in activity recognition explored aspects of quality there is still no common understanding in the community as to what defines the quality of an activity and particularly what is "high" or "low" quality.

The term "quality" has been widely discussed in other fields, such as management research. The International Standards Association defines quality as the "degree to which a set of inherent characteristics fulfils requirements" [27] and Crosby [10] defines it as "conformance to specifications". What these definitions have in common is the fact that one starts with a product specification and a quality inspector measures the adherence of the final product to this specification. These definitions make it clear that in order to measure quality, a benchmark is needed to measure the quality of a product against, in this case its product specification. Adapting this idea to the qualitative activity recognition domain it becomes clear that if we can *specify* how an activity has to be performed we can measure the quality by comparing its execution against this specification.

From this, we define quality as *the adherence of the execution of an activity to its specification.* From this, we define a qualitative activity recognition system as *a software artefact that observes the user's execution of an activity and compares it to a specification.* Hence, even if there is not a single accepted way of performing an activity, if a manner of execution is specified, we can measure quality.

4. QUALITATIVE ACTIVITY RECOGNITION

Based on the definition of quality and qualitative activity recognition it is worth discussing which are its main aspects and challenges. Qualitative activity recognition differs from conventional activity recognition in a distinctive way. While the latter is concerned with recognising *which* activity is performed, the former is concerned with assessing *how (well)* it is performed. Once an activity is specified, the system is able to detect mistakes and provide feedback to the user on how to correct these mistakes.

This directly raises three important questions. First, is it possible to **detect mistakes** in the execution of the activity? Traditional activity recognition has extensively explored how to classify different activities. Will these methods work as well for qualitative assessment of activities? The second question is how we **specify activities**. Two approaches are commonly used in activity recognition: a sensor-oriented approach, in which a classification algorithm is trained on the execution of activities and a model-oriented approach, in which activities are represented by a human skeleton model. The third is how to **provide feedback** in real-time to improve the quality of execution. Depending on how fast the system can make the assessment, the feedback will either be provided in real-time or as soon as the activity is completed. Real-time feedback has the advantage of allowing the user to correct his movements on the go, while an offline system might make use of more complex algorithms and provide useful information without distracting the user.

In this work, we try to tackle each aspect separately. In the next sections we explore a wearable sensor-oriented classification approach for the detection of mistakes, we describe a model-oriented approach to the specification of activities and we evaluate two feedback systems implemented using the modelling approach.

5. DETECTION OF MISTAKES

The goal of our first experiment was to assess whether we could detect mistakes in weight-lifting exercises by using activity recognition techniques. we recorded users performing the same activity correctly and with a set of common mistakes with wearable sensors and used machine learning to classify each mistake. This way, we used the training data as the activity specification and the classification algorithm as the means to compare the execution to the specification.

For data recording we used four 9 degrees of freedom Razor inertial measurement units (IMU), which provide three-axes acceleration, gyroscope and magnetometer data at a joint sampling rate of 45 Hz. Each IMU also featured a Bluetooth module to stream the recorded data to a notebook running the Context Recognition Network Toolbox [3]. We mounted the sensors in the users' glove, armband, lumbar belt and dumbbell (see Figure 1). We designed the tracking system to be as unobtrusive as possible, as these are all equipmentm commonly used by weight lifters.

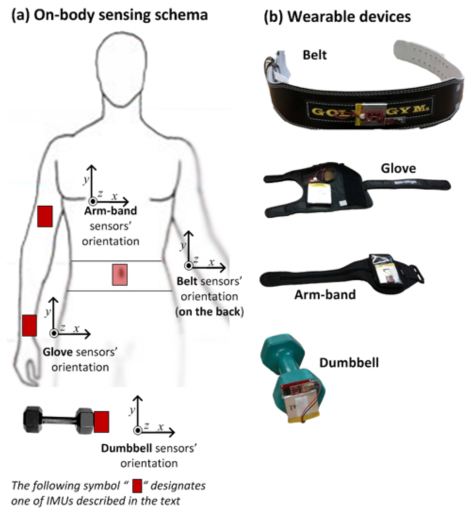

Figure 1: Sensing setup

Participants were asked to perform one set of 10 repetitions of the Unilateral Dumbbell Biceps Curl in five different fashions: exactly according to the specification (Class A), throwing the elbows to the front (Class B), lifting the dumbbell only halfway (Class C), lowering the dumbbell only halfway (Class D) and throwing the hips to the front (Class E). Class A corresponds to the specified execution of the exercise, while the other 4 classes correspond to common mistakes. Participants were supervised by an experienced weight lifter to make sure the execution complied to the manner they were supposed to simulate. The exercises were performed by six male participants aged between 20-28 years, with little weight lifting experience. We made sure that all participants could easily simulate the mistakes in a safe and controlled manner by using a relatively light dumbbell (1.25kg).

5.1 Feature extraction and selection

For feature extraction we used a sliding window approach with different lengths from 0.5 second to 2.5 seconds, with 0.5 second overlap. In each step of the sliding window approach we calculated features on the Euler angles (roll, pitch and yaw), as well as the raw accelerometer, gyroscope and magnetometer readings. For the Euler angles of each of the four sensors we calculated eight features: mean, variance, standard deviation, max, min, amplitude, kurtosis and skewness, generating in total 96 derived feature sets.

In order to identify the most relevant features we used the feature selection algorithm based on correlation proposed by Hall [14]. The algorithm was configured to use a "Best First" strategy based on backtracking. 17 features were selected: in the belt, were selected the mean and variance of the roll, maximum, range and variance of the accelerometer vector, variance of the gyro and variance of the magnetometer. In

Table 1: **Recognition performance**

Window Size	FPR	Recall	AUC	Precision
0.5s	3.9	85.0	97.4	84.9
1.0s	1.8	93.5	99.5	93.5
1.5s	1.0	96.5	99.8	96.5
2.0s	0.7	97.2	99.9	97.2
2.5s	0.5	98.2	99.9	**98.2**

the arm, the variance of the accelerometer vector and the maximum and minimum of the magnetometer were selected. In the dumbbell, the selected features were the maximum of the acceleration, variance of the gyro and maximum and minimum of the magnetometer, while in the glove, the sum of the pitch and the maximum and minimum of the gyro were selected.

5.2 Recognition Performance

Because of the characteristic noise in the sensor data, we used a Random Forest approach [28]. This algorithm is characterized by a subset of features, selected in a random and independent manner with the same distribution for each of the trees in the forest. To improve recognition performance we used an ensemble of classifiers using the "Bagging" method [6]. We used 10 random forests and each forest was implemented with 10 trees. The classifier was tested with 10-fold cross-validation and different windows sizes, all of them with 0.5s overlapping (except the window with 0.5s). The best window size found for this classification task was of 2.5s and the overall recognition performance was of 98.03% (see Table 1). The table shows false positive rate (FPR), precision, recall, as well as area under the curve (AUC) averaged for each of the 5 tested on 10-fold cross-validation over all 6 participants (5 classes). With the 2.5s window size, the detailed accuracy by class was of: (A) 97.6%, (B) 97.3%, (C) 98.2%, (D) 98.1%, (E) 99.1%, (98.2% weighted average).

We also used the leave-one-subject-out test in order to measure whether our classifier trained for some subjects is still useful for a new subject. The overall recognition performance in this test was 78.2 %. The result can be attributed to the small size of the datasets (approx 1800 instances each dataset, extracted from 39.200 readings on the IMUs), the number of subjects (6 young men), and the difficulty in differentiating variations of the same exercise, which is a challenge in Qualitative Assessment Activities. The use of this approach requires a lot of data from several subjects, in order to reach a result that can be generalized for a new user without the need of training the classifier. The confusion matrix of the leave-one-subject-out test is illustrated on Figure 2.

5.3 Conclusion

The advantage of this approach is that no formal specification is necessary, but even though our results point out that it is possible to detect mistakes by classification, this approach is hardly scalable. It would be infeasible to record all possible mistakes for each exercise. Moreover, even if this

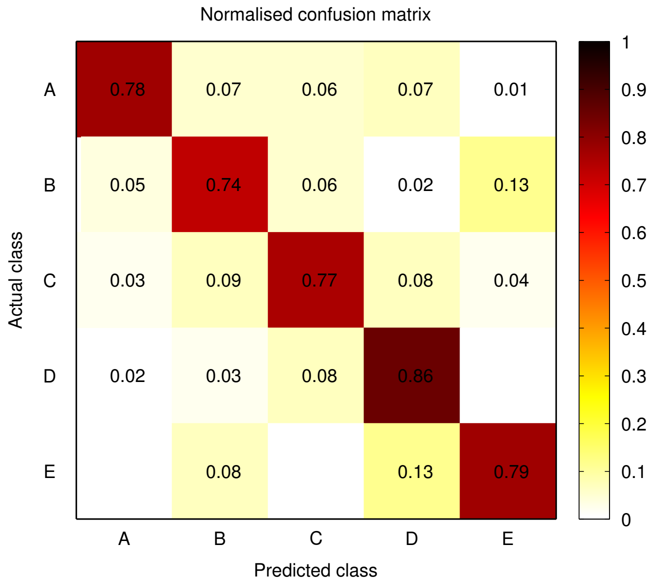

Figure 2: Summed confusion matrix averaged over all participants and normalised across ground truth rows.

was possible, the more classes that need to be considered the harder the classification problem becomes.

6. MODEL-BASED SPECIFICATION

Due to the inherent problems of the classification approach, we concentrated our efforts into trying to formalize a way of specifying activities and recognizing mistakes by looking at deviations from the model in the execution. This section outlines our approach to qualitative activity recognition systems for weight lifting that helps minimize the effort of translating specifications into systems. We implemented a C# framework for the development of such applications using the Microsoft Kinect sensor for body motion tracking. We illustrate our approach on the example of building a feedback system for the Unilateral Dumbbell Biceps Curl and the Unilateral Lateral Dumbbell Raise exercises using our framework.

6.1 Activity Selection

An activity must have an appropriate granularity to be analysed. If the activity is too complex, it is more appropriate to break it down into smaller activities. In our example, even though a weight lifting exercise is commonly performed in sets of 6-12 repetitions, for our purposes we consider an activity as a repetition of the exercise. This way we can analyse each repetition separately. A Biceps Curl repetition involves raising and lowering the dumbbell, so we define the beginning of the activity as when the user starts to lift it and the end as when it reaches the initial position again.

6.2 Activity Specification

The activity should be specified as clearly as possible in natural language. The clearer the specification is the easier it will be to model the activity. In our example, we used as the specification the instructions provided by a weight lifting book [7]. An activity specification can be comprised

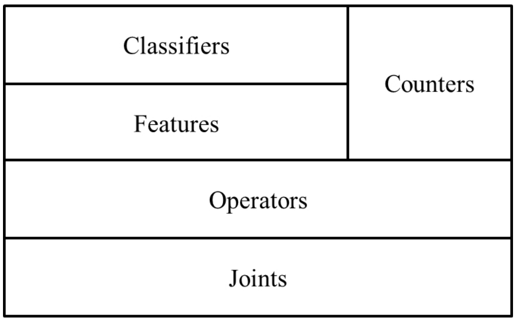

Figure 3: Layered architecture of our model, that receives as input the raw position of the joints as provided by a tracking system and outputs a class of quality for the activity.

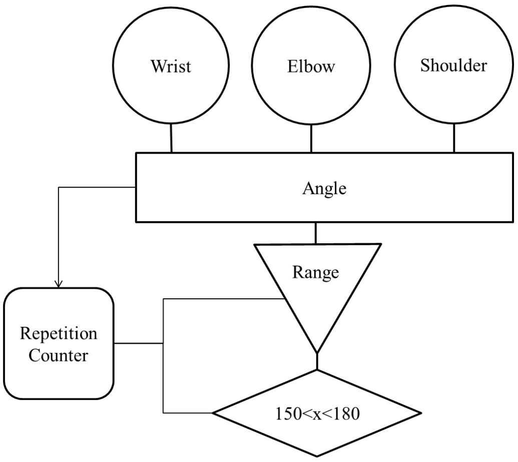

Figure 4: Model for instruction 4. From the joints' position coordinates, we extract the angle between them and its range to count repetitions. For each repetition, we calculate the overall range and check whether it is within the specified limits.

of several instructions. For the Unilateral Dumbbell Biceps Curl, the specification we used, adapted from [7], was the following: (1) Stand solidly upright; (2) Your feet should be shoulder-width apart; (3) Your shoulders should be down; (4) Curl the dumbbell in an upward arc. Curl the dumbbell to the top of the movement when your biceps is fully contracted; (5) Elbows pointing directly down and return to the start position; (6) Don't lean back and throw your hips to the front.

6.3 Activity Modelling

From each instruction in the activity specification we create a model of the recognition mechanism according to the components in the framework. The components can be of five different classes: **Joint**, **Operator**, **Feature**, **Counter** and **Classifier**. The model architecture is illustrated in Figure 3 and Figure 4 shows an example of an instruction modelled accordingly.

A **Joint** in our model is the XYZ position of each of the 20 joints provided by the Microsoft Kinect 1.0 Beta2 SDK. Different instructions will make use of different sets of Joints. In the example in Figure 4, the joints are the Wrist, the Elbow and the Shoulder of each side.

Operators represent operations performed on top of the raw position coordinates of a single joint or a set of joints. The implemented operators include the XYZ coordinates, distance and angles between joints. For example, in the modelling of Instruction 4, we could describe the movement in terms of the trajectory of the hand, but this wouldn't be ideal because it would depend on the length of each user's arms. Hence, we use an Operator to convert it to the angle between the Hand, Elbow and Shoulder instead, because this is not a user-dependent measure. **Feature** components buffer the data that is provided by the operators and perform statistical analysis (such as mean, standard variation, range, energy, etc.) on a dataset when an event is triggered. In the example, because we want to make sure the movement is complete, we measure the range of the angle.

The classification is triggered by **Counters**. In our approach, we can classify an exercise in two ways: continuously (with features being sampled in short intervals) or discretely (with features being sampled after every repetition). If you need a feature to be monitored after a specified time interval, you can use the **Clock Counter**. If you want the feature to be extracted for each repetition, you can use a **Repetition Counter**, which triggers events after detecting a repetition. Finally, the classification of the quality of the execution of the instruction is performed by **Classifier** components. These can range from performing simple thresholding operations to running more complex machine learning algorithms. In Figure 4, the Angle between the Wrist, Elbow and Shoulder is fed into the Repetition Counter, that uses a strong filter and a peak counting algorithm to detect repetitions. When a new repetition is detected, this component trigger the calculation of the range.

Once the model is complete, the class library we implemented allows the programmer to translate directly the components in the model into an object-oriented application. All that is required is to input the parameters in the instantiation of the components and to connect the components by subscribing to each other's events. We modelled and implemented the feedback systems for 3 exercises: Unilateral Dumbbell Biceps Curl, Unilateral Dumbbell Triceps Extension and Unilateral Dumbbell Lateral Raise.

6.4 Parameter Adjustment

There will be times when the available instruction is more qualitative than quantitative, so some instructions should be adjusted and parameterised to account for that. For example, one of the instructions for the Biceps Curl was to "Curl the dumbbell in an upwards arc towards your shoulder". This instruction does not provide the metrics to unambiguously build the model. One possible interpretation is: the angle between the wrist, the elbow and the shoulder should go from 180 to 0 degrees. However, these values need to be tested and adjusted to make sure they correspond to the measurements provided by the Kinect SDK. The framework allows you to debug this step using events that lets you monitor each step of the analysis.

IWe tried to keep the Classifiers as simple as possible so they could be easily tweaked on the spot. This is useful for a real life scenario where the trainer might want make minor alterations in the specification. For example, a general specification for the Biceps Curl says that one should curl the Dumbbell all the way to the top. However, it is possible that the trainer might want the athlete to perform the exercise only halfway to the top in order to stimulate specific muscle fibres. Our system is prepared to allow these parameter modifications to be made easily.

6.5 User Feedback

In the user interface, the system should give feedback for the conformance to each one of the instructions in the specification separately. The feedback should be as clear as possible using different visual and auditory cues. The classifiers output different classes of quality that can be translated into traffic lights that would turn green if the specification was OK and red in case of problems in the exercise, for example. Because of the complex nature of the exercises, it is also recommended to give feedback on how to improve the technique.

7. PROVIDING USER FEEDBACK

Besides mistake recognition and activity specification the third and last aspect of qualitative activity recognition that we explore in this paper is the feedback to the user. We carried out a user study to evaluate a system developed using our framework to check whether our approach to qualitative activity recognition can lead to improvement in the quality of exercise performance. First, participants were asked to fill in a questionnaire regarding their experience with weight lifting prior to the execution of the exercises. The 8 participants were all male, 20-28 years old, with little or no experience in weight lifting. The feedback systems include a traffic light that indicates whether an instruction is being performed correctly and messages instructing the user on how to improve the execution. They also featured a range of motion indicator and a repetition counter. The user could see himself performing the exercise, with the feed from the camera built in the Kinect sensor. The interface is illustrated in Figure 5. Then the participants were asked to perform the Unilateral Biceps Curl and the Unilateral Lateral Raise. We wanted to compare the execution of these exercises with and without the feedback system, so we provided them with a written description of the expected execution of the exercise and asked them to perform each exercise with a hand while the feedback system was turned off. Then, we turned the feedback system on and asked them to perform the same exercises but now with another hand, in order to minimize the effects of tiredness. Each exercise was performed in three sets of ten repetitions with increasing weights of 1.25kg, 3.0kg and 7.0 kg. Participants were instructed to stop whenever they felt uncomfortable. We recorded data using a Kinect sensor connected to a Windows 7 PC. The feedback was provided using a 27-inch LCD display. After the exercises, participants were asked to fill in another questionnaire regarding the experience with the feedback system.

7.1 Results

With the Lateral Raise feedback system users made 23.48% fewer mistakes per repetition, while with the Biceps Curl

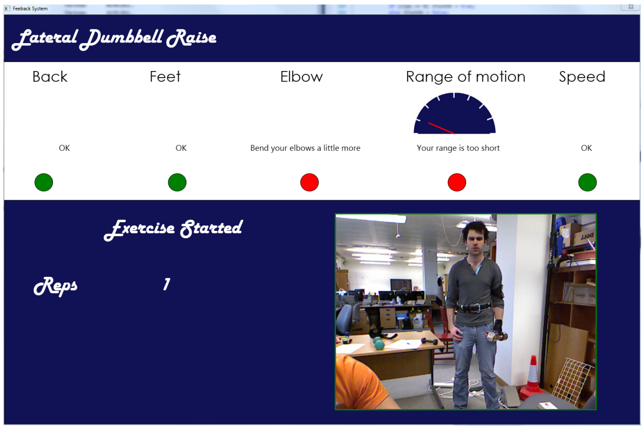

Figure 5: User interface and feedback system for the Unilateral Dumbbell Lateral Raise exercise.

Table 2: Questionnaire results.

Question	Mean	Std
How helpful do you think such system is in a gym environment?	4.57	0.53
How clear was the presentation of information?	4.14	0.90
How much do you believe the feedback influenced your performance?	3.86	0.90
Did you try to change your movements according to the feedback?	4.71	0.49
Did the feedback improve your performance?	3.57	0.79

feedback system users made 79.22% fewer mistakes. Participants were rank ordered by the number of errors in each of the two conditions. A Wilcoxon matched pairs signed ranks test indicated that the number of errors was significantly lower when using feedback ($Mdn = 4.5$) than without using feedback($Mdn = 26.5$), $Z = 2.52$, $p = .008$, $r = .63$. For the lateral curl exercise a Wilcoxon matched pairs signed ranks test indicated no significant difference between the two conditions, $Z = 1.61$, $p = .125$. These results indicate great potential for such systems in correcting mistakes and consequently improving the quality of weight lifting activities. Table 2 shows the mean and standard deviation of the questionnaire results averaged over all participants. Values correspond to responses on a 5-point Likert scale with 5 representing strongly positive and 1 strongly negative answers. User responses were generally positive regarding doing the exercises with the aid of a feedback system. Some suggestions for improvement include making the messages larger and easier to read and trying out different feedback visualizations, like video or 3D animation.

8. DISCUSSION

8.1 Activity Specification

In this paper we tried out two approaches to specifying activities. In the first one, we specified the activity by using machine learning techniques. We recorded data of users performing an exercise in different ways, some of which corresponded to common mistakes made in the execution of this particular activity. In this approach, the activity specification is achieved with the classifier training data.

Even though no formal specification is necessary for this approach, the classifier training can turn out to be a quite tiresome task. In order to train a robust, user-independent classifier, it necessary to record several executions of the same exercise. Moreover, recording all possible mistakes is a potentially infinite task, due to the complexity of human motion and all possible combinations of mistakes. Also, if the specification changes, it is necessary to record the training data all over again.

Due to these problems, we formalised the specification of activities by developing a model that translates instructions in the exercise specification into implementable components. This approach proved to have several advantages. First, because each instruction is specified separately, the system can be tailored for different types of users, by using larger or smaller subsets of instructions. For example, a system for a user with back problems might include instructions specifically designed for monitoring of the spine.

This also allows the reuse of instructions that are common across different exercises. For instance, both exercises we implemented contained an instruction to keep the feet shoulder-width apart and we were able to use the same modelling and implementation for both cases. Moreover, because the system architecture is based in layers, even if the instruction is not exactly across all layers, sometimes it is possible to reuse some of them, with just some parameter changes. For example, both the Biceps Curl and the Lateral Raise require the user to lift the dumbbell. This requires the system to monitor an angle between joints and check for its range. By changing which joints are being monitored and the target range parameters for the exercise, we were able to reuse components. Also, because the model is parameterized, if it is flexbile to specification changes. If the change is small, tweaking some parameters could be enough, but major modifications can be accomplished by swapping some components in the model.

8.2 Mistake Detection

In the classification approach, mistake detection was done by classifying an execution to one of the mistake classes. As stated previously, the main challenge of this approach is scalability. However, we could detect mistakes fairly accurately. In the model-based approach, we detected mistakes by looking in the execution data for deviations from the model. Even though out model supports complex classifiers, we kept our implementation as simple as possible by using threshold-based classifiers. This allowed us to test the implementation and make adjustments to the parameters quite easily. In a real life setting, this would allow the trainer to tailor parameters according to the user's needs. One weakness of our implementation is that we assume the joint positions provided by the Kinect sensor to be accurate, which is not an entirely unreasonable premise as suggested by [11], but could be an issue for high performance athletes. The general approach, however is not tightly coupled to the tracking system, so the system could be enhanced with the use of a more sophisticated tracking system.

8.3 Feedback

Once we implemented the qualitative assessment systems, we evaluated the feedback provided in a case study. Our results showed significant improvement in the Biceps Curl. The Biceps Curl is fairly well known exercise, that people do without taking the time to think about the technique, so the system worked well in aiding users correcting mistakes. Only a small improvement was detected for the Lateral Raise. Even though users made almost a quarter of mistakes made without the system, we can't say that the results are statistically significant. This points out to a potential in the system, but further inspection is necessary. We attribute this result to the difficulty of performing this exercise with the provided weights. The Lateral Raise stimulates mainly the deltoids, which are significantly weaker muscles than the biceps, so a fall in performance was expected.

Users were generally very positive about the system. Some claimed to be "more conscious about movements due to both the camera image and feedback visualisation" and to feel "more confident in the movements I was making and able to correct mistakes." The use of feedback systems was praised: "Without the feedback system you can not be sure whether you are doing the exercise properly", indicating that this field of research deserves more attention.

Some participants thought the simple interface was good ("simple signals gave exact instructions on what to correct") while others had some suggestions on how to improve it ("red and green could be avoided (...) as color blind people will not be able to see the difference" and "I would prefer images that illustrate what to improve"), showing that how to design interfaces for such systems is a challenge on its own.

9. CONCLUSION

In this work we investigated qualitative activity recognition on the example of assessing the quality of execution of weight lifting exercises. We formalized a definition of execution quality and explored three key aspects of qualitative activity recognition, namely how to deal with specifying activities, detecting mistakes and providing feedback.

While the detection of a small number of mistakes is possible using standard pattern recognition techniques, the proposed model-based approach scales better and also allows to encode expert knowledge into the activity specification. The significant positive impact of the real-time feedback provided by our system underlines the potential and opens up the discussion of the wider applicability of this approach to other activities and domains.

10. REFERENCES

[1] A. Baca and P. Kornfeind. Rapid feedback systems for elite sports training. *IEEE Pervasive Computing*, 5(4):70–76, 2006.

[2] M. Bächlin, K. Förster, and G. Tröster. Swimmaster: A wearable assistant for swimmer. In *Ubicomp '09: Proceedings of the 11th international conference on Ubiquitous computing*, pages 215–224, New York, NY, USA, Sept. 2009. ACM.

[3] D. Bannach, P. Lukowicz, and O. Amft. Rapid prototyping of activity recognition applications. *Pervasive Computing, IEEE*, 7(2):22 –31, 2008.

[4] M. Beetz, B. Kirchlechner, and M. Lames. Computerized real-time analysis of football games. *IEEE Pervasive Computing*, 4(3):33–39, 2005.

[5] S. N. Blair. Physical inactivity: the biggest public health problem of the 21st century. *British Journal of Sports Medicine*, 43(1):1–2, 2009.

[6] L. Breiman. Bagging predictors. *Machine Learning*, 24:123–140, 1996. 10.1007/BF00058655.

[7] B. Bridle, editor. *Strength Training: The Complete step-by-step guide to sculpting a stronger body*. Dorling Kinderley Limited, 2011.

[8] F. Buttussi and L. Chittaro. Mopet: A context-aware and user-adaptive wearable system for fitness training. *Artificial Intelligence in Medicine*, 42(2):153–163, 2008.

[9] K.-h. Chang, M. Y. Chen, and J. Canny. Tracking free-weight exercises. In *Proceedings of the 9th ACM International Conference on Ubiquitous computing*, volume 4717, pages 19–37. Springer, 2007.

[10] P. B. Crosby. *Quality is Free*. McGraw-Hill, 1979.

[11] T. Dutta. Evaluation of the kinectŹ sensor for 3-d kinematic measurement in the workplace. *Applied Ergonomics*, 2011.

[12] M. Ermes, J. Parkka, J. Mantyjarvi, and I. Korhonen. Detection of daily activities and sports with wearable sensors in controlled and uncontrolled conditions. *Information Technology in Biomedicine, IEEE Transactions on*, 12(1):20 –26, jan. 2008.

[13] M. Gallagher. Ten most common causes of training injury. *Muscle & Fitness*, June 1996.

[14] M. A. Hall. *Correlation-based Feature Subset Selection for Machine Learning*. PhD thesis, Department of Computer Science, University of Waikato, Hamilton, New Zealand, Apr. 1999.

[15] N. Y. Hammerla, T. Ploetz, P. Andras, and P. Olivier. Assessing motor performance with pca. In *Proceedings of the International Workshop on Frontiers in Activity Recognition using Pervasive Sensing*, 2011.

[16] T. C. Havens, G. L. Alexander, C. Abbott, J. M. Keller, M. Skubic, and M. Rantz. Contour tracking of human exercises. In *Proceedings of Computational Intelligence for Visual Intelligence*, pages 22–28, 2009.

[17] J. Hey and S. Carter. Pervasive computing in sports training. *IEEE Pervasive Computing*, 4(3):54, 2005.

[18] P. Iskandar, Y. Hanum, L. Gilbert, and G. Wills. The design of effective feedback in computer-based sport training. In *Proceedings of the 7th International Symposium on Computer Science in Sport*, pages 1–13, 2009.

[19] Z. Y. Kerr, C. L. Collins, and R. D. Comstock. Epidemiology of weight training-related injuries presenting to united states emergency departments, 1990 to 2007. *The American Journal of Sports Medicine*, 38(4):765–771, 2010.

[20] J. F. Knight, A. Schwirtz, F. Psomadelis, C. Baber, H. W. Bristow, and T. N. Arvanitis. The design of the sensvest. *Personal and Ubiquitous Computing*, 9(1):6–19, 2004.

[21] F. Michahelles and B. Schiele. Sensing and monitoring professional skiers. *IEEE Pervasive Computing*, 4(3):40–46, 2005.

[22] A. Moeller, L. Roalter, S. Diewald, M. Kranz, N. Hammerla, P. Olivier, and T. Ploetz. Gymskill: A personal trainer for physical exercises. In *Proceedings of the International Conference on Pervasive Computing and Communications*, 2012.

[23] S. L. Mullane, N. Chakravorti, P. P. Conway, and A. A. West. Design and implementation of a user-centric swimming performance monitoring tool. *Proceedings of the Institution of Mechanical Engineers, Part P: Journal of Sports Engineering and Technology*, 225(4):213–229, 2011.

[24] G. O'Donovan, A. J. Blazevich, C. Boreham, A. R. Cooper, H. Crank, U. Ekelund, K. R. Fox, P. Gately, B. Giles-Corti, J. M. R. Gill, and et al. The abc of physical activity for health: a consensus statement from the british association of sport and exercise sciences. *Journal of Sports Sciences*, 28(6):573–591, 2010.

[25] N. Oliver and F. Flores-Mangas. Mptrain: a mobile, music and physiology-based personal trainer. In *Proceedings of the 8th Conference on Human-computer interaction with mobile devices and services*, pages 21–28. ACM, 2006.

[26] A. Reiss, G. Hendeby, G. Bleser, and D. Stricker. Activity recognition using biomechanical model based pose estimation. In *Smart Sensing and Context*, volume 6446 of *Lecture Notes in Computer Science*, pages 42–55. Springer, 2010.

[27] I. O. F. Standardisation. Iso 9000:2005 quality management systems – fundamentals and vocabulary. *Management*, 3, 2005.

[28] L. B. Statistics and L. Breiman. Random forests. In *Machine Learning*, pages 5–32, 2001.

[29] J. Staudenmayer, D. Pober, S. Crouter, D. Bassett, and P. Freedson. An artificial neural network to estimate physical activity energy expenditure and identify physical activity type from an accelerometer. *Journal of Applied Physiology*, 107(4):1300–1307, 2009.

[30] C. Strohrmann, H. Harms, G. Tröster, S. Hensler, and R. Müller. Out of the lab and into the woods: Kinematic analysis in running using wearable sensors. In *Proceedings of the 13th ACM International Conference on Ubiquitous Computing (UbiComp 2011)*. ACM, Sept. 2011.

[31] A. Zinnen, U. Blanke, and B. Schiele. An analysis of sensor-oriented vs. model-based activity recognition. In *Wearable Computers, 2009. ISWC '09. International Symposium on*, pages 93 –100, sept. 2009.

[32] A. Zinnen, C. Wojek, and B. Schiele. Multi activity recognition based on bodymodel-derived primitives. In *Proceedings of the 4th International Symposium on Location and Context Awareness*, LoCA '09, pages 1–18, Berlin, Heidelberg, 2009. Springer-Verlag.