Modeling the Model Athlete :
Automatic Coaching of Rowing Technique

Simon Fothergill, Robert Harle, Sean Holden

Computer Laboratory, University of Cambridge,
Cambridgeshire, CB3 0FD, U.K.
{jsf29, rkh23, sbh11}@cam.ac.uk

Abstract. Watching athletes allows coaches to provide both vital feedback on how well they are performing and on ways to improve their technique without causing or aggravating injuries. The thoroughness and accuracy of this traditional observation method are limited by human ability and availability. Supplementing coaches with sensor systems that generate accurate feedback on any technical aspect of the performance gives athletes a fall back if they do not have enough confidence in their coach's assessment.

A system is presented to model the quality of arbitrary aspects of rowing technique found to be inconsistently well performed by a set of novice rowers when using an ergometer. Using only the motion of the handle, tracked using a high-fidelity motion capture system, a coach trains the system with their idea of the skill-level exhibited during each performance, by labeling example trajectories. Misclassification of unseen performances is encouragingly low, even for unknown performers.

Keywords: Novel Applications, Sports coaching, Quality, Rowing Technique, Body motion, Intelligent Sensing Systems, Spatiotemporal Pattern Recognition, Shape Analysis.

1 Introduction

When practising a physical skill, reassurance that our technique is good or feedback on how to correct it, are crucial for motivation and improvement. Learning a technique so we perform correctly without thinking requires lengthy (re)training of "muscle memory" as well as studying the underlying biomechanical principles, especially if the technique is complex or we need to break old habits. Real-time feedback can help athletes when learning what good and bad technique feels like. As they may not be able to feel whether the whole performance is correct and can not stop to observe themselves, feedback is required from a third party such as a coach. In this paper, the authors choose to investigate generating feedback to supplement coaches of the Olympic sport of rowing. The multi-faceted technique requires a very high level of consistency and precision at high speed between multiple rowers. On the water, a few centimeters difference in a stroke can unbalance and decelerate a boat so there is real need for continuous personal feedback. Rowing can be practiced very realistically in a laboratory environment by using a stationary ergometer.

Good coaches will always be in demand but automated analysis of body movements may offer advantages over human coaching: world-class coaches are often busy and expensive. They often coach in squads to instill competition, giving each athlete a fraction of their time during which they might give unrepresentative and damaging feedback. Even experienced coaches can not simultaneously concentrate on all aspects of a technique, can be biased towards athletes they know, can fail to notice differences in a performer they are used to watching and can become impaired by factors such as fatigue.

An automated, surrogate system could analyze athletes continuously whenever they train, objectively focusing on any number of aspects simultaneously. As motion capture systems continue to develop, this hardware will become much more affordable than a real person. On-athlete sensors could also offer a more accurate perspective from inside the boat rather than from a distance. When athletes are interested in semantically sophisticated descriptions of a performance, informal discussions between coaches and the authors showed they may be less comfortable representing a performance as digitized trajectories of points, rather than watching their moving flesh. By automating analysis of the trajectories, athletes could instead be presented with comprehendible and familiar ontological elements from the domain of rowing, which could also be used as automatically constructed, low bandwidth indices into recorded performances.

The authors have circumstantial evidence from the sports science community that biomechanical theory can not always justify coaches' behavior and that the latter find difficultly in explaining their judgments. Supervised learning provides a framework for directly associating coach-level feedback with performances recorded using a few sensibly, but not precisely placed sensors, whereas a traditional (bio)-mechanical approach would require explicit and precise formulation of rules for every facet of the technique.

The physical sensation of rowing drives the terminology to include words such as "relaxed", "fluid", "too", "sufficient" or "that looks right!" Their amorphous and complex nature is hard to capture using explicit biomechanical formulae, especially when the authors tried to use some coaches' explanations. Approximations may have detrimental or unobvious effects on classification performance through including or missing out certain factors; when rowing even small movements are important.

Measurements of performances of a similar quality still have inherent noise from the sensors and differences in where they are worn. Characterization of what technique works best for different athletes also adds to variation that is not suited to exact mechanical formulae. Even for more tangible aspects such as "overreaching" or "leaning back too far", rules would need to be personalized and require expert biomechanical advice to be trustworthy. The authors found that a large number of biomechanical measurements are typically taken and instrumenting athletes is time consuming. Rules that rely on specific sensors become dependable on those sensors functioning. By using such rules the system is customized to the domain of rowing and the same system could not be taught different sporting techniques.

The highly complex nature of rowing where large intra-class variations are present but important differences between classes are made due to multiple, occasionally unobvious aspects changing by large or very small amounts, makes this an apt challenging problem for machine learning, especially when using a limited set of body movements. How well a learning system might recognize the quality of an individual aspect of technique when performed by different people and with different levels of skill, whilst ignoring other changing aspects of the technique, is a question that has not been addressed for the domain of rowing as far as the authors are aware.

The process of coaching athletes includes observation and analysis of their technique to determine the quality of their performance, deciding what encouragement or corrections are necessary and formulating how to communicate this so that athletes appreciate it. This involves explaining and demonstrating correct technique whilst the athletes rest so they can be prompted to recall it as necessary during a later performance. The quality of a performance can be described at various levels of precision from an overall judgment of the whole performance, to the quality of individual aspects of the technique, to a highly detailed description of each muscle. This work focuses on determining the quality of individual aspects of rowing technique using binary classification as this provides a simple yet informative representation of common judgments.

2 Previous Work

Much research exists into the capture and analysis of human body motion, including recognition of everyday activities to specific gesticulations. However, judgment of the quality of a physical performance is less common and work that exists tends to focus on a few direct biomechanical measurements that are somehow related to quality within the medical domain.

Work in [2] uses HMMs to recognize a vocabulary of gestures that could make up a surgical procedure. Using different observations of the performance, a naïve judgment of skill is based simply on the number of different gestures used in a performance, or the percentage of time spent on each gesture. No rigorous validation is completed. Authors of [5] admit that the objective measurements of surgeons' dexterity as recorded by a motion capture system within an operating theatre does not capture the complexity and differences in styles involved in assessing the skill of different surgeons. The authors of [4, 10] use HMMs to obtain values for overall quality of a complete surgical performance, rather than individual aspects. The method used in [10] splits the performances into sub-sections with semantic significance and used the distance from a template performance as a measure of quality. The authors of [4] calculated the log-likelihood of a performance by applying one model to the whole performance, disregarding any internal structure and achieved a recognition of performance that correlated highly with an expert (r = 0.93 p<0.001).

Recognition of sporting activities is the focus of work in [3], where elements of a technique such as the tennis serve are recognized in video footage. However this is done by high-level scene analysis. The work of [1] evaluates overall quality in Karate using many markers and requiring models of prototype performances that are morphed to new performances. Results are presented for one person performing 20 different Karate moves. Recognition accuracy varies over technique but is on average 85%. The author of [6] discussed the need for individual aspects of a performance to be assessed, but proposed that an automated rule-based system would struggle with the flexible or fuzzy nature of the existing rubrics.

3 Description of the System

The system provides binary classification of a population of strokes taken from a number of different performances. Strokes are separated based on the quality of an arbitrary aspect of the rowing technique used. It is able to learn the judgment rule from a set of labeled performances and provide validation of how well it generalizes over an unseen population of strokes.

3.1 Motion Capture

The VICON motion capture system [11] was used with 10 cameras and default parameter values to track the position of single points on multiple objects augmented with 3 or more retro-reflective markers (see Fig. 1). The sample rate was set to 200Hz. The objects are automatically identified using unique markers topologies.

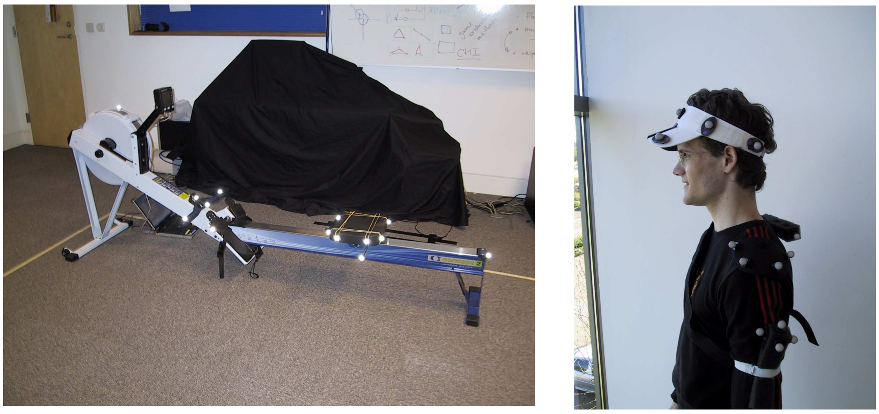

Fig. 1. Ergometer and rower augmented with VICON markers. The erg, its seat and its handle were 3 objects. Parts of the body were captured for future use.

Some markers on body parts and the seat were occluded for short periods of time no longer than half a second, which is approximately a quarter of the period of a stroke. Conventional video was also used to record the performances for review later.

3.2 Preprocessing

Occluded portions of the trajectories are recovered using linear interpolation. All trajectories are transformed from the motion capture co-ordinate system to an erg-coordinate system (see Fig. 2) to make analysis of the data more intuitive.

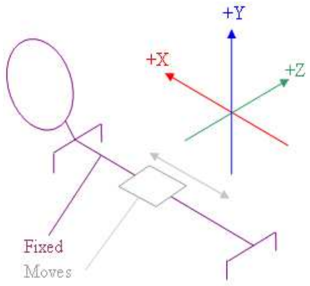

Fig. 2. Erg co-ordinate system

Markers of the seat were used to define the X-axis as the seat moves along it, and the X-Y plane from which the normal gives the Z-axis. The seat exhibited a measured deviation of less than one mm from the averaged X-axis. This work only uses markers on the handle, seat and erg-frame. This trajectory is segmented into separate strokes using an algorithm that detects the main troughs in the handle's X-coordinate. The first and last strokes were discarded in case they were unrepresentative as the athletes were warming up/down.

3.3 Feature Extraction

The tempo-spatial features listed below are used to represent the handle trajectory for a stroke. The stroke rate (number of strokes every minute) is made invariant as it is taken to be independent of the aspects of technique used in this study.

- Distance travelled
- Speed moments
- Shape smoothness
- Length
- Wobble
- Ratio
- Height
- Drive smoothness
- Drive angle
- Shape moments
- Recovery smoothness
- Recovery angle

Abstract Features. The trajectory's distance was computed, as was the length (along X-axis) and height (along Y-axis). Moments of the 2D shape formed by orthogonally projecting the trajectory onto the X-Y plane were also computed, using the projected co-ordinates for each sample s, as values $h_x(s)$ and $h_y(s)$ and summing over all samples s, from the trajectory. The following were computed for shape: λ^{11}, λ^{12}, λ^{21}, λ^{02}, λ^{20} (1). The instantaneous, mean-subtracted speed at each sample point s, was used as the function $\psi(s)$ to compute the speed moments: μ^{11}, μ^{12}, μ^{21}, μ^{02}, μ^{20} (1).

$$\lambda^{pq} = \sum (h_x(s)^p\, h_y(s)^q) \,, \qquad \mu^{pq} = \sum (h_x(s)^p\, h_y(s)^q\, \psi(s)) \,, \qquad (\mathbf{1})$$

Physical Performance Features. The following were chosen based on preliminary experiments to investigate how well they disambiguated a small set of performances of differing overall quality. Wobble (lateral variance) was computed as the variance in the distance each trajectory sample is from the best fit line when they are projected into the X-Z plane. Smoothness of the drive and recovery in time were computed by low-pass filtering the mean-subtracted, instantaneous speed of the handle at 3Hz (as the raw motion data was too noisy), identifying large accelerations by taking the second derivative and summing the absolute values of this signal. Smoothness of the shape was computed by forming a signal from the distances of each subsequent trajectory sample point, to the centroid of the trajectory and low-pass filtering at 6Hz, followed by measuring the flatness of the absolute value of the second differential of this signal by counting the number of samples that are less than 0.4 ms^{-2}. All smoothness measures were normalized to the length of the stroke.

Rowing Features. Ratio is the ratio of the time spent on the recovery to the time spent on the slide. Drive and recovery angles are the angles between the best fit line and X axis when the points are projected into the X-Z plane.

3.4 Supervised Learning

Normalization and Negation. Each feature used in the model has parameters computed to adjust its values to be roughly within the same range as all other features in order to weight and compare them more easily. This is done for each stroke by subtracting the minimum and dividing by the difference between the maximum and minimum values of the feature, in the non-normalized training set. If the values from the training set for a single feature correlate highly but negatively with each strokes' labeled scores, the normalized feature's complement is used instead of the original feature, i.e. the original feature value is subtracted from 1. The labeled score for each stroke is normalized to 0 (bad) and 1 (good).

Classifiers. Each stroke from a population is automatically scored using a linear combination of a weighted bias and the weighted features. The weights of the model are learnt using one of two different methods: The first solves the system of simultaneous equations formed from the feature vectors of each stroke as rows of a matrix multiplied by a column vector of weights, which is made equal to a corresponding column vector of labeled scores for each stroke. It is solved using the Moore-Penrose pseudo-inverse of the feature matrix. The second method uses gradient descent to adjust the weights, which are initially set to zero after each iteration through the whole training set. A learning rate of 0.001 is used for 750 iterations to minimize the sum of the square of the differences between the labeled and machine scores of the training set. The training is repeated for a number of times using different training sets formed by leaving out a different set of strokes from the whole original population. These unseen stroke(s) are classified using the trained model and their scores recorded. Each stroke is left out exactly once. When considering multiple performers all strokes from one performer are left out at once. This "leave-one-out" validation (whether it be one stroke or one athlete) gives machine scores for each stroke in the population, ensuring it hasn't been used to train the classifier that scores it. The scores are summarized using Pearson's coefficient for the correlation between the coach and the machine scores and by using the percentage of strokes misclassified. The latter is computed by thresholding the two classes for the machine scores at a point which minimizes the misclassification. This value is calculated exhaustively.

Sensitivity Analysis. For the second training method the optimum number of iterations is chosen by repeatedly calculating the percentage of misclassified strokes, having trained using a different number of iterations, increasing by 50 each time from 50 to 750. For both methods, an optimum feature set is chosen that minimizes the percentage of misclassified strokes by repeating the sensitivity analysis for the number of iterations, for different feature sets. Each set is formed using exhaustive backwards-selection by removing the least important features until only 1 is left. The initial set is all the features and importance is measured by summing the weights over all training sets when using the optimum number of iterations. The class threshold, number of iterations and feature set that give the smallest misclassification error are assumed when presenting the results for different represented populations of strokes.

4 Empirical Validation of System

These experiments gave an indication of the system's capabilities in real-world coaching scenarios. Each stroke was scored by using the score a single amateur coach gave to the whole performance it was from and by reviewing the videos the coach considers at least 95% of the strokes of a performance to be representative of the quality they assigned to the whole performance. The coach (lead author) has been rowing for 2 years and was confident of assessing basic technique in novice rowers. Strokes by different rowers are differentiated as a person's physique and skill level effect the set of strokes recorded as does their conscious decisions for how to row and how much attention to pay the coach. Six novice, male rowers in their mid-twenties, between 60kg and 90kg were used with very little or no rowing experience. They were not initially fatigued and rowed at a comfortable rate in an uncontrived manner. The amount of input from the coach during a performance affects the consistency of a the technique during a performance and can increase how much of the stroke must be ignored or generalized over when only interested in scoring one aspect.

Each rower was given a basic explanation of how to row and gave an initial performance. A number of things were usually wrong with this performance so a coaching process was then repeated until no obvious faults existed or the rower was exhausted: the coach evaluated the last performance and taught the rower how to improve an aspect. The rower then gave another performance during which the coach helped them to maintain the improved technique for the increasing number of corrected aspects. Each performance lasted for approximately 30 strokes and the corrected aspects were maintained for at least 95% of the strokes.

The final features chosen for each experiment are recorded to investigate their usefulness. For either training method all features are used in at least 60% of the final features sets. Shape and speed moments λ^{02}, λ^{20} μ^{02} and μ^{20} were the only four used in at least 90% of the final feature sets for both algorithms.

4.1 Coaching Single Aspects for Individuals

Coaches often try to improve a single aspect of technique at a time in order to simplify the training process. This requires the coach to judge the quality of this aspect and issue prompts to the rower when necessary as they try to correct only this aspect. As only one aspect was focused on for each consecutive performance during the data collection, the system effectiveness at this task is shown by using pairs of consecutive performances from each novice as sample populations, see Table 1.

Table 1. Percentages of misclassified strokes for all experiments when coaching single and multiple aspects of the technique of individual rowers.

Rower	Coached aspect (chronological order)	Moore-Penrose training		Gradient Descent training	
		Single aspect	All aspects	Single aspect	All aspects
1	Separate arms/legs	0	0	0	0
	Overreaching	0	0	0	1
2	Separate arms/legs	0	-	3	-
3	Separate arms/legs	0	0	0	0
	Overreaching	0	0	2	1
4	Overreaching	0	0	0	0
	Shins vertical	0	0	0	0
	Early open back	0	0	2	1
5	Leaning back	0	3	0	6
	Quick hands	0	4	0	7
	Rushing slide	0	2	0	6
	Early open back	3	0	5	0
6	Overreaching	0	0	0	0
	Separate arms/legs	0	0	1	3
	Quick hands	0	0	1	1

4.2 Coaching Multiple Aspects Simultaneously for Individuals

It can be more efficient to issue coaching calls on multiple aspects depending on what the rower does wrong. This requires a coach to make judgments about one aspect, no matter what the qualities of other aspects are. As data were collected on each rower whilst several aspects of the stroke were performed at different qualities, realistic combinations of different qualities for several aspects were observed as were realistic amounts of variation in strokes with the same level of quality for one aspect. Using all performances from each individual, results of the system judging the quality of each coached aspect for strokes where multiple aspects are changing are shown in Table 1.

4.3 Coaching Single Aspects across Different Individuals

In order to coach a new athlete who has not been seen before the system can only base a decision of how well they row on performances of other people. Being able to do this would provide more evidence of being able to recognize the quality of an aspect when other parts of the stroke are performed differently as in §4.2. By having enough examples, representative populations of strokes can be formed using consecutive pairs of performances form multiple people, all of whom are improving the same aspect. Table 2 shows the recognition results and the number of rowers used for each aspect.

Table 2. Percentages of misclassified strokes for aspects across a number of different athletes.

Rowers	Aspect	Moore-Penrose training	Gradient Descent training
2	Quick hands	9	5
2	Early open back	33	29
3	Separate arms/legs	21	21
4	Overreaching	12	12

5 Discussion, Conclusions and Further Work

Analysis of the final feature sets shows it is wise to make use of all the features. The common choice of the four moment features suggests they may capture fundamental and important characteristics of quality of technique and analysis of relative weights would confirm their suitability as good candidates for recognizing other sports.

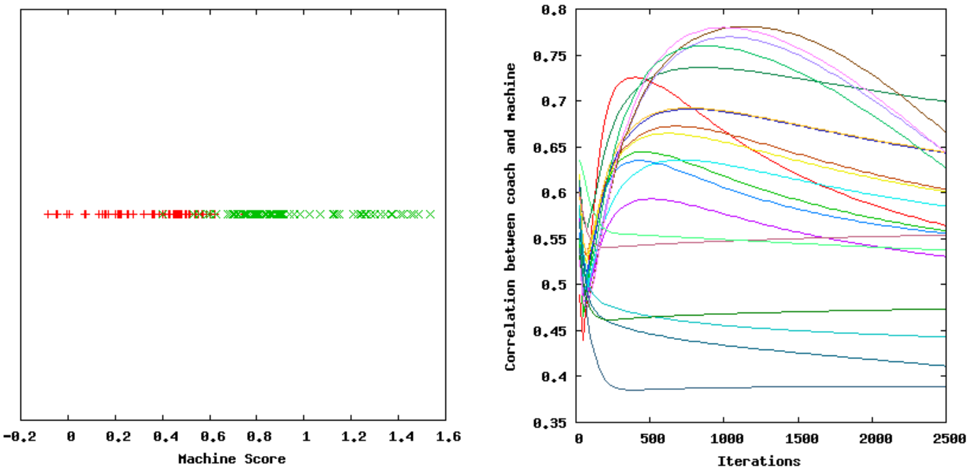

Fig. 3. Judging strokes from unseen people for the aspect “Quick Hands”: machine score (left), correlation values using gradient descent up to 2500 iterations, for different feature sets (right).

Given the small amount of evidence for coaching higher numbers of aspects, when coaching up to four, for an single athletes, very low misclassification was found. The Moore-Penrose solution gave slightly better results. The minimum possible training error does minimize the generalized misclassification results and gradient descent does not find it in time at this learning rate. Recognition of strokes from unseen people gave quite low (5-21%) misclassification for three of the experiments. The fourth training set was only one person. The gradient descent method was slightly better this time: Tightening the model to fit the training data too closely by using Moore-Penrose had adverse effects on the generalization to unseen performances as shown in Fig. 3 where too many iterations caused the highest correlating feature sets to correlate less well, even though the training error was observed never to increase. Inter-variation of data from different athletes is observed to be generally greater than an athlete's intra-variation, suggesting why the results are significantly worse for unseen athletes. Further work including scoring each stroke separately, using more markers and estimating the weights differently or using other widespread classifiers. Semantically sophisticated analysis of physical performance would be possible if ontological elements from a domain of sports techniques were identifiable. Encouraging classification results are achieved for qualitative assessment of rowing technique but the limits of the recognition algorithms are not yet fully characterized.

Acknowledgments. Thanks are due to EPSRC, U.K. and the Digital Technology and Rainbow groups and Professor Andy Hopper at the Computer Laboratory, University of Cambridge U.K. for funding, support, equipment and volunteers.

References

1. Ilg, Mezger & Giese. Estimation of Skill Levels in Sports Based on Hierarchical Spatio-Temporal Correspondences. DAGM 2003, LNCS 2781, pp. 523-531, 2003.
2. Murphy, Vignes, Yuh, Okamura. Automatic Motion Recognition and Skill Evaluation for Dynamic Tasks. EuroHaptics 2003, 2003.
3. Zhong & Chang. Structure Analysis of Sports Video Using Domain Models. International Conference on Multimedia & Expo, Tokyo, Japan, August, 2001, pp. 920—9233, 2001
4. Leong, Nicolaou, Atallah, Mylonas, Darzi & Yang. HMM Assessment of Quality of Movement Trajectory in Laparoscopic Surgery. MICCAI 2006, LNCS 4190, pp. 752-759, 2006
5. Dosis, Aggarwal, Bello, Moorthy, Munz, Gillies & Darzi. Synchronized Video and Motion Analysis for the Assessment of Procedures in the Operating Theater. Arch Surg. 140, pp. 293-299, 2005.
6. Gordon. Automated Video Assessment of Human Performance. J. Greer (ed) Proceedings of AI-ED 95. pp. 541-546, 1995.
7. Australian Capital Territory Rowing Association, Rowing Technique for Coaches, http://www.rowingact.org.au/SDO/TECHNIQUE_1.html
8. Concept2, Rowing Technique, Faults and Corrections http://www.concept2.co.uk/training/faults_corrections.php.
9. el Kaliouby & Robinson, Generalization of a vision-based computational model of mind-reading. In: First International Conference on Affective Computing and Intelligent Interaction, 2005.
10. Rosen, Solazzo, Hannaford & Sinanan. Objective Laparoscopic Skills Assessments of Surgical Residents Using Hidden Markov Models Based on Haptic Information and Tool/Tissue Interactions. The Ninth Conference on Medicine Meets Virtual Reality, 2001.
11. VICON, a motion capture system, http://www.vicon.com