

Abstract

Humans have the desire to achieve better performance and outcome of their efforts. In performances such as sports, yoga or professions such as surgery, humans strive to attain perfection and efficiency. They improve their performances by comparing themselves with others or seeking feedback from experts. In this regard the question arises here; can we make a human action assessment system which can support the users with a flexibility of using the system at preferred time and place, thereby avoiding the dependency on a trainer.

The key challenge in action quality assessment is the lack of an appropriate metric as it is quite a subjective task with a significant influence of human (expert) bias. The labels like scores or skill levels provided by human judges lack interpretability. Thus, it will be interesting to develop an automated action assessment system that can bring more interpretability and objectivity to this domain. Action assessment system can be made more objective by considering a set of reference high-precision/expert-level action videos. In this thesis, we propose techniques to assess human actions, where the problem of action assessment is transformed into the problem of comparing a given action video with a reference video. We consider two use-cases : Yoga and Sports, and assess actions like Sun Salutation, Diving and Gymnastic Vaults.

We contribute a new Sun Salutation Assessment dataset that includes expert and non-expert performances and their respective ground truth judgments. Sun Salutation is a long term complex action with many possible errors and is a good test-bed for human action assessment techniques. Physical exercises like Sun Salutation and Aerobics are repetitive in nature. Such exercises require the performer to achieve consistent dynamic poses and smooth transitions. The quality check for such exercises involves the analysis of attributes such as pace, consistency, smoothness, etc. of the performers. We propose a framework to identify jerky and inconsistent movements in a performance that involves action segmentation using Hidden Markov Model, followed by an inter-pose timing analysis.

Based on the performance skill, humans can be divided into experts, mid-level, and amateurs. Amateur performers are prone to miss or wrongly perform a part of some action. We propose a template matching-based approach where the pose sequence of the test performer is compared with an expert sequence and the action segments that were missed or anomalously performed are reported. Template-based matching is a fairly simple approach to identify differences and has difficulty in scaling when the number of templates tends to grow. In practice, actions can be performed correctly in multiples ways. Thus there is a need for comparison of the test sequence to all expert renderings possible for an action sequence. Towards this, we propose an unsupervised sequence-to-sequence autoencoder-based model that learns to reconstruct all expert videos. The skill level of a test performance is judged based on how well the learned model can reconstruct the test sequence. The closer the test performance is to an expert, the more accurately it gets reconstructed.

Actions like diving or gymnastic vaults are constrained with limited availability of expert performances, which makes it difficult to train an autoencoder-based assessment model. To address this limitation, we propose a Deep Metric Learning based framework, where a Long Short Term Memory (LSTM)-based Siamese network learns to predict if two videos in a pair are similar or dissimilar based on the difference in their scores. The learned model is utilized for action scoring where the performances are compared with the reference expert performance to determine the score.

...