ViziCal: Accurate Energy Expenditure Prediction for Playing Exergames

Miran Kim[1] mirank@cse.unr.edu
Jeff Angermann[2] jeff@medicine.unr.edu
George Bebis[1] bebis@cse.unr.edu
Eelke Folmer[1] efolmer@cse.unr.edu

[1]Department of Computer Science and Engineering
[2]School of Community Health Sciences
University of Nevada, Reno

ABSTRACT
In recent years, exercise games have been criticized for not being able to engage their players into levels of physical activity that are high enough to yield health benefits. A major challenge in the design of exergames, however, is that it is difficult to assess the amount of physical activity an exergame yields due to limitations of existing techniques to assess energy expenditure of exergaming activities. With recent advances in commercial depth sensing technology to accurately track players' motions in 3D, we present a technique called Vizical that uses a non-linear regression approach to accurately predict energy expenditure in real-time. Vizical may allow for creating exergames that can report energy expenditure while playing, and whose intensity can be adjusted in real-time to stimulate larger health benefits.

Author Keywords
Machine Learning, Computer Vision, Exergames, Gestures, Exercise, Health, Energy Expenditure, Physical Activity.

ACM Classification Keywords
I.2.10 Vision and Scene Understanding: Motion

INTRODUCTION
Recent studies show that short bouts of high-intensity training, such as the popular CrossFit workouts, have great potential to significantly improve fitness levels [32]. Though the durations are shorter than typical aerobic activities, the benefits are usually longer lasting and the improvements to cardiovascular health and weight loss are more significant [5]. These findings are particularly interesting in the context and design of *exergames*, e.g., video games that use exertion-based interfaces to promote physical activity, fitness, and gross motor skill development [26]. As exergames typically involve short bouts of upper and whole body gestures, they may be considered as a form of high-intensity training and could yield similar associated health benefits. As video games are considered powerful motivators, many researchers have argued that exergames could be an important tool in fighting the current childhood obesity epidemic [29].

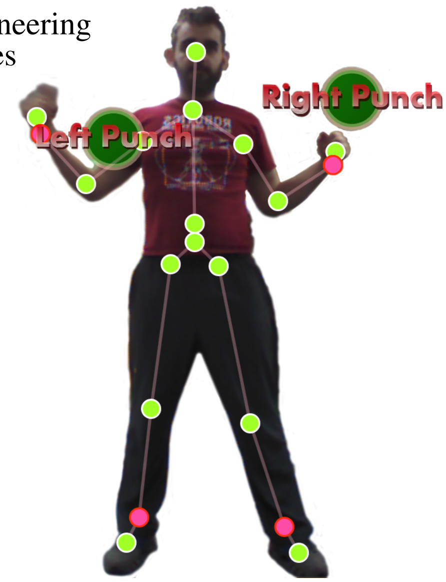

Figure 1: Recent commercially available depth sensing cameras, such as Microsoft Kinect, allow for accurately tracking skeletal joint positions of a user playing an exergame.

Recent studies reveal that popular exergames, such as Nintendo Wii, stimulate greater energy expenditure (EE) than when playing sedentary video games [14, 22], but they don't achieve levels of physical activity that are comparable with conventional types of exercise [15, 10] or the sports they simulate [13]. Though exergames have been introduced in some physical education programs [28], interest in using them beyond entertainment purposes seems to be waning [2].

A challenge in the design of exercise games is that it is difficult to measure the amount of EE an exergame yields. This is due to a number of limitations of existing methods for measuring EE, for example, the use of heart rate is limited as it can be influenced by psychological factors, such as excitement. Wearable activity monitors predominantly measure locomotion and are limited in being able to capture upper body motions, which are often used in exercise games. EE can be more accurately measured using calorimetric techniques such as metabolic gas analysis systems, but these techniques are expensive and require a significant amount of training.

We present Vizical, a low-cost and easy to use technique for accurately assessing the EE of exergames. Vizical takes advantage of recent advances in motion sensing input controller technology (Microsoft Kinect) to accurately track a player's movements in 3D (See Figure 1). Vizical may be used to design exergames that: (1) stimulate higher levels of physical activity, (2) report current or total EE; and (3) can adjust their intensity in real-time.

Submitted for review to UIST 2013.

BACKGROUND AND RELATED WORK

Total energy expenditure (TEE) of humans has three components: (1) basal metabolic rate (BMR), thermic effect of food (TEF); and the energy expenditure of activity (AEE). A complete review of all different types of energy expenditure assessment techniques is beyond the scope of this paper, though a comprehensive overview is provided by Levine [24]). In general, EE can be measured using three different approaches. We will discuss the most popular methods for each approach and discuss their limitations in being able to assess EE of exergames.

Direct calorimetry methods measures the rate of heat loss from a subject using a calorimeter. A calorimeter is typically implemented in the form of an isothermal or adiabatic chamber. Because these systems are extremely expensive to build and require enormous expertise to operate, these techniques are typically only used in highly specialized labs [24].

Indirect calorimetry methods measures oxygen consumption and/or CO2 production and convert this into EE using various formulas. Various techniques exist, for example, total collection systems, such as the Douglas bag [33] can be used to collect expired air from a subject wearing a mask. Confinement systems, such as a respiratory chamber, have the subject contained in gas-tight room. For both approaches, oxygen consumption can be measured by changes in volume and/or composition of the expired air. These systems are very accurate but not portable. Also, they do require extensive training to use the systems. Because the response time to observe changes in oxygen consumption is typically high (>10 minutes), these systems are typically used to measure EE over longer periods of time. Different from the above mentioned systems, metabolic gas analysis systems are considered to be more practical and preferable; the systems require subjects to inhale air through a mask and expire via a non-return valve in order to measure oxygen consumption (VO2) and/or CO2 production (VCO2). In recent years, portable VO2 metabolic systems have been developed [23]. Such portable systems operate untethered and transmit data wirelessly for various physical activities outside a laboratory setting. These systems are very accurate and have fast response times (< 30s), but they are expensive and do require extensive training. Additionally, subjects may experience respiratory discomfort using metabolic systems while exercising.

Non-calorimetric methods predict EE by extrapolating physiological or kinematic measurements. These methods are often calibrated against calorimetric techniques [24]. Because heart rate typically increases when users engage in physical activity, it can be used as a proxy for physical activity [12]. Heart rate monitors are portable, low-cost, and unobtrusive, and they can be used over longer periods of time. Heart rate is subject to large individual variation and is a poor proxy for exertion for children due to developmental considerations [8]. The most significant limitation of this method in context of exergames is that heart rate is affected by psychological factors, such as excitement. This is problematic as playing (non-active) video games have been found to elevate heart rate [16], without users actually getting physically active. The energy cost of physical activities can also be assessed using kinematic measurements, typically in the form of wearable accelerometers. Approaches for EE estimation using wearable accelerometers can be classified in two categories: (1) physical-based [17], and (2) regression-based [9] approaches. Physical-based approaches rely on a model of the human body, where velocity or position information is estimated from accelerometer data and kinetic motion and/or segmental body mass is used for estimating EE. Regression-based approaches, on the other hand, estimate EE by directly mapping accelerometer data to EE. The simplest regression-based approach is to estimate EE from a single accelerometer placed at the hip using linear regression. This approach has been extended to using non-linear regression models (i.e., to fully capture the complex relationship between acceleration and EE) [36] and multiple accelerometers (i.e., to account for upper or lower body motion which is hard to capture from a single accelerometer placed at the hip) [30, 37]. Combining accelerometers with heart rate monitors [30] has been shown to improve EE estimation significantly.

In the context of exergames, a limitation of using accelerometers is in their ability to capture total activity, as accelerometers only selectively record movement of the part of the body to which they are attached. Accelerometers worn on the hip are primarily suitable for gait or step approximation, but they cannot capture upper body movement; if worn on the wrist, locomotion is not accurately recorded. Exergames typically involve whole body motions, such as punches and kicks. Increasing the number of accelerometers will increase the accuracy of capturing total body movement, but it is often inconvenient and not practical, due to cost. Another limitation is that accelerometers can't report EE in real time and they have a limited sensitivity (e.g., due to band pass filters they only detect accelerations between 0.05 and 2G, which prevents them from detecting light activities or discriminating between levels of vigorous activities [7]).

Exergaming studies. A recent analysis [3] of exergaming studies reveals that most exergames are only able to engage children into light-to-moderate levels of physical activity with games involving whole body gestures stimulating larger amounts of EE than games that only involve upper body gestures. The Center for Disease Control recommends children should engage into 60 minutes of moderate-to-vigorous levels of activity daily, including 20 minutes of vigorous activity [1]. Given these criteria, current exergames do not stimulate types and amounts of physical activity required to maintain cardiorespiratory fitness. The majority of these studies used heart rate and accelerometers to predict EE. The only two studies [38, 25] that found moderate levels of physical activity for whole body exergames, used VO2 metabolic systems to measure EE. These findings seem to support our claim that the recent exergaming criticisms are to an extent unfounded, as techniques with significant limitations are being used to assess EE of exergames.

Related work. Four unique approaches closely related to ViziCal were identified. Krohn and Boisclair assessed the metabolic cost of fish swimming in an aquarium by tracking their motions using a stereo camera while measuring their total oxygen consumption [21]. Image analysis was used to measure the swimming speed of each fish, which was subsequently used as a feature in predicting EE. This technique is not well suited for assessment of human activity, as the continuous motion of forced swimming is not applicable to the rapid and sporadic movements typically associated with exergaming. In another approach utilizing video analysis, Kaneko et al. assessed EE in spacelab astronauts performing gymnastic exercises during spaceflight [18]; this approach is not applicable to our problem since the astronaut's motions are significantly different from how our exergames are played. Ogsnach et al. employed video analysis to model and calculate EE in professional soccer players [27], but their experimental design only accounted for running performance. Botton et al. estimated EE in tennis using computer vision [4]. Their approach classified tennis EE as one of five different types of activities, i.e., walking, running, returning, serving, and sitting down. Each of these activities has a different amount of EE associated with it, which was measured using a portable VO2 metabolic system. Overall EE was then calculated by sequencing the activity using video analysis. As this approach was limited to activity segregation and classification, it does not approximate the continuous range of motion movement accounted for in our model.

VIZICAL

We have developed a non-calorimetric technique called Vizical that can predict EE of exergaming activities using the rich amount of kinematic information acquired using a commercially available 3D camera (Kinect). Kinect is a controllerless input device used for playing video games and exercise games for the Xbox 360 platform. This sensor can track up to six humans in an area of 6m^2 by projecting a speckle pattern onto the users body using an IR laser projector. A 3D map of the users body is then created in real-time by measuring deformations in the reference speckle pattern. A single depth image allows for extracting the 3D position of 20 skeletal joints at 200 frames per second. This method is invariant to pose, body shape and clothing [34]. The joints include hip center, spine, shoulder center, head, shoulder, elbow, wrist, hand, hip, knee, ankle, and foot (See Figure 1). The estimated joint locations include the direction that the person is facing, which allows for distinguishing between the left and right joints for shoulder, elbow, wrist, hand, hip, knee, ankle and foot. Studies have investigated the accuracy of Kinect [19, 11], which found that the depth measurement error ranges from a few millimeters at the minimum range (70cm) up to about 4 cm at the maximum range of the sensor (6.0m).

We adopt a regression based approach by directly mapping kinematic data collected using the Kinect to EE, since this has shown good results without requiring a model of the human body [30, 37]. The EE of playing an exergame is acquired using a portable VO2 metabolic system, which provides the ground truth for training a regression model (see Figure 2). Given a reasonable amount of training data, the regression model can then predict EE of exergaming activities based on kinematic data captured using a Kinect sensor

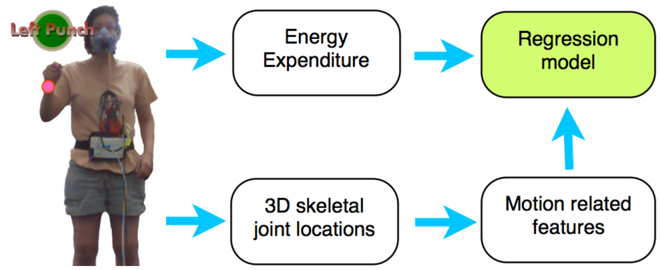

Figure 2: Kinematic information and EE of a subject playing an exergame is obtained using a portable VO2 metabolic system. From the skeletal joint location data, various motion related features are extracted to train a regression model using the collected ground truth data.

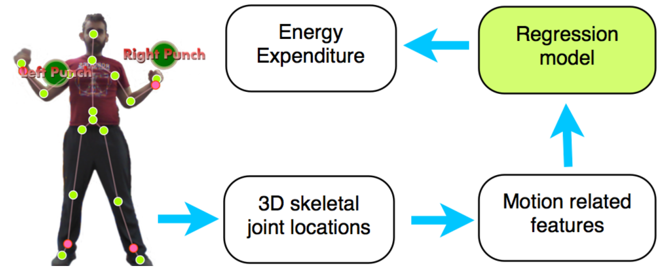

Figure 3: Based on kinematic information the regression model can then predict the EE of an activity.

(see Figure 3). Accelerometer based approaches typically estimate EE using a linear regression model over a sliding window of one-minute length using the number of acceleration counts per minute [9] (e.g., the sum of the absolute values of the acceleration). A recent study found several limitations for linear regression models to accurately predict EE using accelerometers [20]. This study further suggests nonlinear regression models may be able to better predict EE associated with upper body motions and high-intensity activities.

For ViziCal, we employ Support Vector Regression (SVR)[35], a popular regression technique that has good generalizability and robustness against outliers and supports non-linear regression models. SVR can approximate complex non-linear relationships using kernel transformations. Kinect allows for recording human motion at a much higher spatial and temporal resolution. Where accelerometer based approaches are limited to using up to five accelerometers simultaneously, ViziCal can take advantage of having location information of 20 joints. This allows for detecting motions of body parts that do not have attached accelerometers such as the elbow or the head. Though accelerometers sample at 32Hz, they report accumulated acceleration data in 1 second epochs. Their sensitivity is also limited (0.05 to 2 G). Because ViziCal acquires 3D joint locations at 200Hz, accelerations can be calculated more accurately and with a higher frequency. Besides using acceleration, features from more powerful, view-invariant, spatial representation schemes of human motion can be used, such as histograms of 3D joints [40]. Besides more accurate EE assessment, ViziCal has a number of other benefits: (1) Accelerometers can only be read out using an external reader, where ViziCal can predict EE in real time, which may allow for real-time adjustment of the intensity of an exergame; (2) Subjects are not required to wear any sensors, though they must stay within range of the Kinect sensor; and (3) Accelerometers typically cost several hundreds of dollars per unit whereas a Kinect sensor retails for $150.

EXPERIMENT

An experiment was conducted to demonstrate the feasibility of Vizical to accurately predict the EE of playing an exergame. This experiment provides insight into the following two research questions: (1) What type of features are most useful in predicting EE? (2) What is the accuracy of ViziCal compared with accelerometer based approaches?

Instrumentation

For our experiment, the Kinect for Windows sensor is used, which offers improved skeletal tracking over the Kinect for Xbox 360 sensor. Though studies have investigated the accuracy of Kinect, these were limited to non-moving objects [19, 11]. We measured the accuracy of the Kinect to track moving joints using an optical 3D motion tracking system with a tracking accuracy of 1mm. We anticipated the arms to be most difficult to track, due to their size; therefore, we attached a marker at the wrist of subjects, close to wrist joints in the Kinect skeletal model. A number of preliminary experiments with two subjects performing various motions with their arms found an average tracking error of less than 10 mm, which we deem acceptable for our experiments. EE is collected using a Cosmed K4b2 portable gas analysis system, which measures pulmonary gas exchange with an accuracy of $\pm 0.02\%$ (O_2), $\pm 0.01\%$ (CO_2) and has a response time of 120ms. This system reports EE in Metabolic Equivalent of Task (MET); a physiological measure expressing the energy cost of physical activities. METs can be converted to calories by measuring an individual's resting metabolic rate.

An exergame was developed using the Kinect SDK 1.5 and which involves destroying virtual targets rendered in front of an image of the player using whole body gestures (See Figure 1 for a screenshot). This game is modeled after popular exergames, such as EyeToy:Kinetic and Kinect Adventures. A recent criticism of exergames is that they only engage their players in light and not vigorous levels of physical activity, where moderate-to-vigorous levels of physical activity are required daily to maintain adequate health and fitness [1]. To allow for ViziCal to distinguish between light and vigorous exergames, a light and a vigorous mode was implemented in our game. The intensity level of any physical activity is considered vigorous if it is greater than 6 METs and light if it is below 3 METs. Using the light mode, players destroy targets using upper body gestures, such as punches, but also using head-butts. We included gestures with the head, as this type of motion is difficult to measure using accelerometers, as they are typically only attached to each limb. We play tested this version with the portable metabolic system using a number of subjects to verify that the average amount of EE was below 3 METs. For the vigorous mode, destroying targets using kicks were added, as previous studies show that exergames involving whole body gesture stimulate larger amounts of EE than exergames that only involve upper body gestures [3]. After extensive play testing, jumps were added to assure the average amount of EE of this mode was over 6 METs. A target is first rendered using a green circle with a radius of 50 pixels. The target stays green for 1 second before turning yellow and then disappears after 1 second. The player scores 5 points if the target is destroyed when green and 1 when yellow as to motivate players to destroy targets as fast as possible. A jump target is rendered as a green line. A sound is played when each target is successfully destroyed. For collision detection, each target can only be destroyed by one specific joint (e.g., wrists, ankles, head). A text is displayed indicating how each target needs to be destroyed, e.g.,"Left Punch"(see Figure 1).

An initial calibration phase determines the length and position of the player's arms. Targets for the kicks and punches are generated at an arm's length distance from the player to stimulate the largest amount of physical activity without having the player move from their position in front of the sensor. Targets for the punches are generated at arm's length at the height of the shoulder joints with a random offset in the XY plane. Targets for the head-butts are generated at the distance of the player's elbows from their shoulders at the height of the head. Jumps are indicated using a yellow line where the players have to jump 25% of the distance between the ankle and the knee. Up to two targets are generated every 2 seconds. The sequence of targets in each mode is generated pseudo-randomly with some fixed probabilities for light (left punch:36%, right punch:36%, two punches:18%, head-butt:10%) and for the vigorous mode (kick:27%, jump:41%, punch:18%, kick+punch:8%, head-butt:5%). Targets are generated such that the same target is not selected sequentially. All variables were determined through extensive play testing as to assure the desired METs were achieved for each mode. While playing the game the Kinect records the subject's 20 joint positions in a log file every 50 milliseconds.

Participants

Previous work on EE estimation [31] has shown that subject independent EE estimation is more difficult than subject dependent estimation. This is because commonly employed regression models fail to account for physiological differences between subject data used to train and test the regression model. For this experiment, we are primarily interested in identifying those features that are most useful in predicting EE. EE will vary due to physiological features, such as gender and gross phenotype. To minimize potential inter-individual variation in EE, which will allow us to concentrate on identifying those features most useful in predicting EE; we collected data from a homogeneous healthy group of subjects, which would allow us to concentrate on The following criteria were used: (1) male; (2) body mass index less than 25; (3) body fat percentage less than 17.5%; (4) age between 18 and 25; (5) exercise at least three times a week for 1 hour. Subjects were recruited through flyers at the local campus sports facilities. Prior to participation, subjects were asked to fill in a health questionnaire to screen out any subjects who met the inclusion criteria but for whom we anticipated a greater risk to participate in the trial due to cardiac conditions or high blood pressure. During the intake, subjects' height, weight and body fat were measured using standard anthropomorphic techniques to assure subjects met the inclusion criteria. Fat percentage was acquired using a body fat scale. A total of 9 males were recruited (average age 20.7 (SD=2.24), weight 74.2 kg (SD=9.81), BMI 23.70 (SD=1.14), fat % 14.41 (SD=1.93)). Our number of subjects is comparable with related regression based studies [39, 36]. Subjects were paid $20 to participate.

Data Collection

User studies took place in an exercise lab. Subjects were asked to bring and wear exercise clothing during the trial. Before each trial the portable VO2 metabolic system was calibrated for volumetric flow using a 3.0L calibrated gas syringe, and the CO2 and O2 sensors were calibrated using a standard gas mixture of O2:16% and CO2:5% according to the manufacturer's instructions. Subjects were equipped with the portable metabolic system, which they wore using a belt around their waist. Also they were equipped with a mask using a head strap where we ensured the mask fit tightly and no air leaked out. Subjects were also equipped with five Actical accelerometers: one on each wrist, ankle and hip to allow for a comparison between techniques. Prior to each trial, accelerometers were calibrated using the subject's height, weight and age. We assured there was no occlusion and that subjects were placed at the recommended distance (2m) from the Kinect sensor. Subjects were instructed what the goal of the game was, i.e., score as many points as possible within the time frame by hitting targets as fast as possible using the right gesture for each target. For each trial, subjects would first play the light mode of the game for 10 minutes. Subjects then rested for 10 minutes upon which they would play the vigorous mode for 10 minutes. This order minimizes any interference effects, e.g., the light bout didn't exert subjects to such an extent that it is detrimental to their performance for the vigorous bout. We limit our data collection to ten minutes, as we consider exergaming activities to be anaerobic and for this experiment are not interested in predicting aerobic activities.

Training the Regression Model

Separate regression models are trained for light and vigorous activities as to predict METs, though all data is used to train a single classifier for classifying physical activities. Eventually when more data is collected, a single regression model can be trained, but for now, the collected data represents disjunct data sets. An SVM classifier was used to classify an exergaming activity into being light or vigorous; only kinematic data and EE for such types of activities was collected. We implemented our classifier and regression models using the LibSVM library. Using the collected ground truth, different regression models are trained as to identify which features or combinations of features yield the best performance. Using the skeletal joint data obtained, two different types of motion-related features are extracted: (1) Acceleration of skeletal joints; and (2) Spatial information of skeletal joints.

Acceleration: acceleration information of skeletal joints is used to predict the physical intensity of playing exergames.

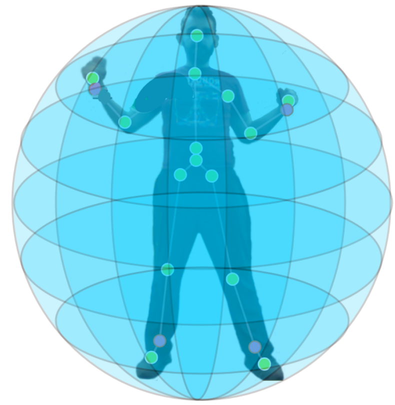

Figure 6: A visual representation of the sphere and its partitioning into bins for the joint binning process.

From the obtained displacement data of skeletal joints, the individual joint's acceleration is calculated in 50ms blocks, which is then averaged over one-minute intervals. We choose to partition our data in one-minute blocks as to allow for comparison with the METs predicted by the accelerometers. Though the Kinect sensor and the Cosmed portable metabolic system can sample with a much higher frequency, using smaller time windows won't allow for suppressing the noise, which exists in the sampled data. There is a significant amount of correlation between accelerations of joints (e.g., when the hand joint moves, the wrist and elbow often move as well as they are linked). To avoid over-fitting our regression model, the redundancy in the kinematic data is reduced using Principal Component Analysis (PCA) where we end up selecting five acceleration features that preserve 90% of the information for the light and 92% for the vigorous model. PCA was applied only because our vectors were very large and we wanted to optimize the performance of training our SVR. We verified experimentally that applying PCA did not affect prediction performance significantly.

Spatial: to use joint locations as a feature, we employ a view-invariant representation scheme called joint location binning [40]. Unlike acceleration, joint binning can capture specific gestures, but it cannot discriminate between vigorous and less vigorous gestures. As acceleration already captures this, we explored joint binning as a complementary feature to improve performance. Joint binning works as follows: 3D space is partitioned in n bins using a spherical coordinate system with an azimuth (θ) and a polar angle (ϕ) that is centered at the subject's hip and surrounds the subject's skeletal model (see figure 6). The parameters for partitioning the sphere and the number of bins that yielded the best performance for each regression model were determined experimentally. For light, the best performance was achieved using 36 bins where θ and ϕ were partitioned into 6 bins each. For vigorous, 36 bins were used where θ was partitioned into 12 bins and ϕ into 3 bins. Binning information for each joint is managed by a histogram with 36 bins; with a total of 20 histograms for all joints were used as a feature vector. Histograms of bin frequencies are created by mapping the 20 joints to appropriate bin locations over one-minute time interval with a 50 ms sampling rate. When bin frequencies are added, the selected bin

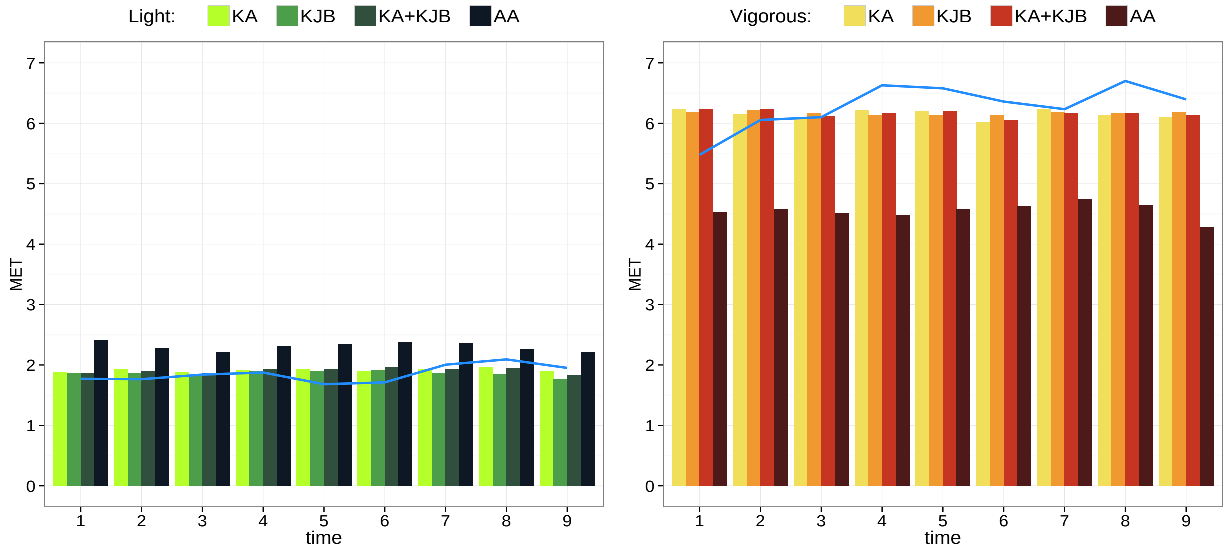

Figure 4: Predicted METs versus ground truth. Figures show predicted METs in one-minute intervals for light (left) and vigorous (right) using three different regression models with each model using the following features: KA uses acceleration; KJB uses joint position; and KA+KJB uses both acceleration and joint position. AA shows predicted MET using the wearable accelerometers. MET's are averaged over 9 subjects. The blue lines show the ground truth MET collected using the portable metabolic system.

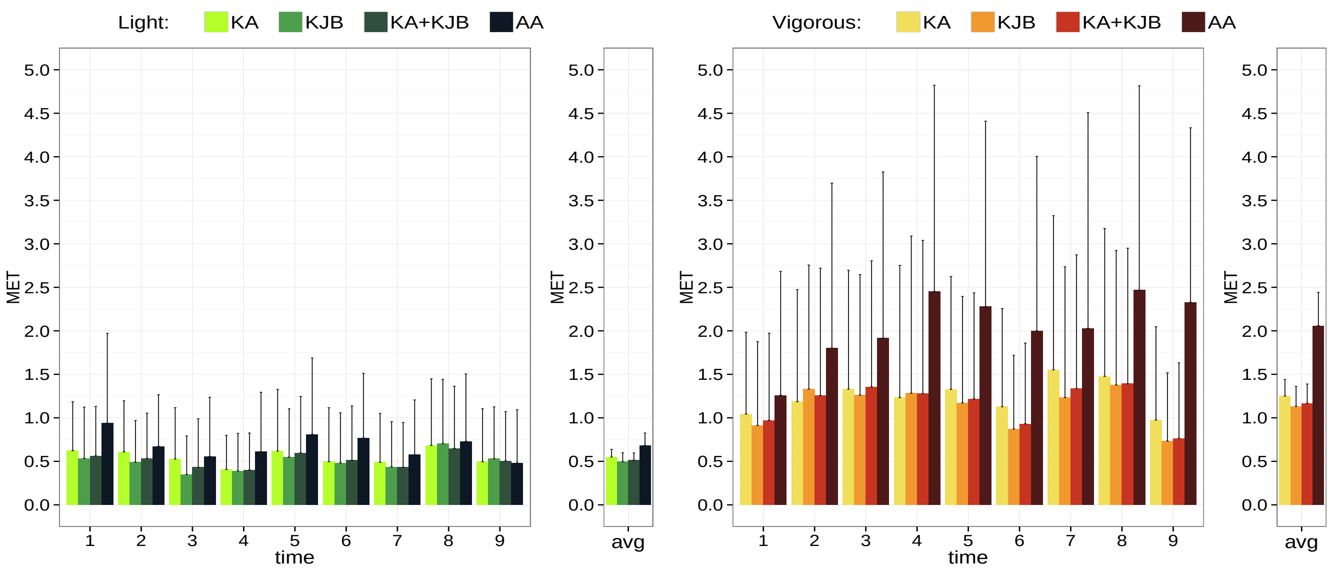

Figure 5: Root mean square (RMS) error of predicted MET versus ground truth for each technique and their averages (see Fig 5 for the legend). Error bars display standard deviation of the RMS error between subjects.

and its neighbors get votes weighted linearly based on the distance of the joint to the center of the bin it is in. To reduce data redundancy and to extract dominant features from the 20 histograms, PCA is used to extract five features retaining 86% of information for light and 92% for the vigorous activities. As the subject starts playing the exergame, it takes some time for their metabolism and heart rate to increase; therefore the first minute of collected data is excluded from our regression model. A leave-one-out approach was used to test the regression models, where data from eight subjects was used for training and the remaining one for testing. We repeat this process so that each subject was used once to test the regression model.

Results

Figure 4 shows the predicted METs of the light and vigorous regression models using three sets of features: (1) acceleration (KA); (2) joint position (KJB) and (3) both (KA+KJB). For the accelerometers (AA), METs are calculated by averaging the METs of each one of the five accelerometers used according to manufacturer's specifications. METs are predicted for each subject and then averaged over the nine subjects; METs are reported in one-minute increments. On average the METs predicted by our regression models are within 17% of the ground truth for light and within 7% for vigorous, where accelerometers overestimate METs with 24% for the light and underestimate METs with 28% for vigorous. These results confirm our assumption that accelerometers predict

EE of exergames poorly. We calculate the root mean square (RMS) error as a measure of accuracy for each technique (see Figure 5). A significant variance in RMS error between subjects can be observed due to physiological differences between subjects. Because the intensity for each exergame is the same throughout the trial, we average METs over the nine-minute trial and compare the performance of all techniques using RMS [39]. For the light exergame, a repeated-measures ANOVA with a Greenhouse-Geisser correction found no statistically significant difference in RMS between any of the techniques ($\boldsymbol{F}_{1.314,10.511} = 3.173$, $p = .097$). For the vigorous exergame, using the same ANOVA, a statistically significant difference was found ($\boldsymbol{F}_{1.256,10.044} = 23.964$, $p < .05$, partial $\eta^2 = .750$). Post-hoc analysis with a Bonferroni adjustment revealed a statistically significant difference between MET predicted by all regression techniques and the accelerometers ($p < .05$). Between the regression models, no significant difference in RMS between the different feature sets was found ($p = .011$).

Classifying Exergame Intensity

To be able to answer the question whether an exergame engages a player into light or vigorous physical activity, we also trained an SVM using all the data collected in our experiment. A total of 162 data points were used for training and testing with each data point containing one-minute of averaged accelerations for each of the 20 joints. Using 9-fold cross-validation we achieved an accuracy of 100%. Once an activity is classified, the corresponding regression model can be used to accurately predict the associated METs.

DISCUSSION AND FUTURE WORK

For vigorous exergaming activities ViziCal predicts MET more accurately than accelerometer-based approaches. This increase in accuracy may be explained by an increase in spatial resolution that allows for capturing gestures, such as head-butts more accurately, and the ability to calculate features more precisely due to a higher sampling frequency. The increase in performance has to be put in context, however, as our regression model was trained and tested using a restricted set of gestures, where accelerometers are trained to predict MET for a wide range of motions, which inherently decreases their accuracy.

We anticipated joint binning to be able to outperform joint acceleration, as it allows for better capturing of specific gestures; but our experiment showed no significant difference in RMS error between both features and their combination. Joint binning however, may yield a better performance for exergames that include more sophisticated sequences of gestures, such as sports based exergames. A drawback of using joint binning as a feature is that it restricts predicting MET to a limited set of motions that were used to train the regression model. The histogram for joint binning for an exergame containing only upward punches looks significantly different from the same game that only contains forward punches. The acceleration features for both gestures, however, are very similar. If it can be assumed that their associated EE do not differ significantly, acceleration may be a more robust feature to use, as it will allow for predicting MET for a wide range of similar gestures that only vary in the direction they are performed, with far fewer training examples required than when using joint binning. Because SVM uses acceleration as a feature, it may already be able to classify the intensity of exergames, who use different gestures from the one we used in our experiment; something we will investigate.

The exergame used for training our regression model uses a range of different motions, but it doesn't cover the gamut of gestures typically used in all types of exergames, which vary from emulating sports to dance games with complex step patterns. We also limited the intensity of our exergame for training the regression models to two extremes, light and vigorous, as these are considered criteria for evaluating the health benefits of an exergame. Rather than having to classify an exergame's intensity a priori, a single regression model that can predict MET for all levels of intensity would be more desirable, especially since moderate levels of physical activity are also considered to yield health benefits [1].

Though no difference was found in performance between acceleration and joint position, there are techniques to refine these features. For example, acceleration can be refined by using coefficient of variation, inter-quartile intervals, power spectral density over particular frequencies, kurtosis, and skew [31]. Joint binning can be refined by weighing bins based on the height of the bin or weighing individual joints based on the size of the limb they are attached to. Since the emphasis of this work was on identifying a set of features that would allow us to predict energy expenditure, we did not perform comparisons using different regression models. In future work, we plan to evaluate random forests regressors, which are used by the Kinect [34] and which typically outperform SVR's for relatively low dimensionality problems spaces like ours [6].

A high variance in RMS error between subjects was observed despite our efforts to minimize variation in EE by drawing our subjects from a homogeneous population. Demographic data should be considered to train different regression models to compensate for inter-individual variations. Alternatively we could possibly calibrate the regression result by incorporating demographic information as input to the regression model or correcting the regression estimates to compensate for demographic differences. Since exergames have been advocated as a promising health intervention technique to fight childhood obesity, it is important to collect data from children. There is an opportunity to use the Kinect to automatically identify demographic data, such as gender, age, height and weight, and automatically associate a regression model with it, without subjects having to provide this information in advance. Future work will also investigate whether we can interpolate between regression models in the case that no demographic match can be found for the subject.

CONCLUSION

In recent years, exercise games have been criticized for not yielding levels of physical activity that are high enough to be considered healthy. We present ViziCal, a tool for accurately predicting energy expenditure (EE) of exergaming activities that doesn't require the user to wear any sensors. Experiments show a significant increase in being able to predict EE for exergames with vigorous motions in comparison with a conventional accelerometer based approach. ViziCal can be used to design exercise games that yield greater health benefits; thus addressing some of the current criticisms of exergames.

ACKNOWLEDGMENTS

This work is supported by NSF Grant IIS-1118074.

REFERENCES

1. Center for disease control, how much activity do you need? **http://www.cdc.gov/physicalactivity/everyone/guidelines/children.html**.
2. New york times, 'exergames' don't make children fitter, **http://parenting.blogs.nytimes.com/2012/06/25/exergames-dont-make-children-more-fit/**.
3. Biddiss, E., and Irwin, J. Active video games to promote physical activity in children and youth: a systematic review. *Arch Pediatr Adolesc Med 164*, 7 (Jul 2010), 664–672.
4. Botton, F., Hautier, C., and Eclache, J.-P. Energy expenditure during tennis play: a preliminary video analysis and metabolic model approach. *J Strength Cond Res 25*, 11 (Nov 2011), 3022–3030.
5. Buchan, D. S., Ollis, S., Young, J. D., Thomas, N. E., Cooper, S.-M., Tong, T. K., Nie, J., Malina, R. M., and Baker, J. S. The effects of time and intensity of exercise on novel and established markers of cvd in adolescent youth. *Am J Hum Biol 23*, 4 (2011), 517–526.
6. Caruana, R., Karampatziakis, N., and Yessenalina, A. An empirical evaluation of supervised learning in high dimensions. In *Proceedings of the 25th international conference on Machine learning*, ICML '08, ACM (New York, NY, USA, 2008), 96–103.
7. Chen, K. Y., Janz, K. F., Zhu, W., and Brychta, R. J. Redefining the roles of sensors in objective physical activity monitoring. *Med Sci Sports Exerc 44*, 1 Suppl 1 (Jan 2012), 213–223.
8. Corder, K., Ekelund, U., Steele, R. M., Wareham, N. J., and Brage, S. Assessment of physical activity in youth. *J Appl Physiol 105*, 3 (Sep 2008), 977–987.
9. Crouter, S. E., and Bassett, Jr, D. R. A new 2-regression model for the actical accelerometer. *Br J Sports Med 42*, 3 (Mar 2008), 217–224.
10. Daley, A. J. Can exergaming contribute to improving physical activity levels and health outcomes in children? *Pediatrics 124*, 2 (Aug 2009), 763–771.
11. Dutta, T. Evaluation of the kinect^{TM} sensor for 3-d kinematic measurement in the workplace. *Appl Ergon 43*, 4 (Jul 2012), 645–654.
12. Gondoni, L. A., Titon, A. M., Nibbio, F., Augello, G., Caetani, G., and Liuzzi, A. Heart rate behavior during an exercise stress test in obese patients. *Nutr Metab Cardiovasc Dis 19*, 3 (Mar 2009), 170–176.
13. Graves, L., Stratton, G., Ridgers, N. D., and Cable, N. T. Comparison of energy expenditure in adolescents when playing new generation and sedentary computer games: cross sectional study. *BMJ 335*, 7633 (December 2007), 1282–1284.
14. Graves, L. E. F., Ridgers, N. D., and Stratton, G. The contribution of upper limb and total body movement to adolescents' energy expenditure whilst playing nintendo wii. *Eur J Appl Physiol 104*, 4 (Nov 2008), 617–623.
15. Graves, L. E. F., Ridgers, N. D., Williams, K., Stratton, G., Atkinson, G., and Cable, N. T. The physiological cost and enjoyment of wii fit in adolescents, young adults, and older adults. *J Phys Act Health 7*, 3 (May 2010), 393–401.
16. Irwin, A. R., and Gross, A. Cognitive tempo, violent video games, and aggressive behavior in young boys. 337–350.
17. Jang, Y., Jung, M., Kang, J., and Chan Kim, H. An wearable energy expenditure analysis system based on the 15-channel whole-body segment acceleration measurement. *Conf Proc IEEE Eng Med Biol Soc 4* (2005), 3834–3840.
18. Kaneko, M., Miyatsuji, K., and Tanabe, S. Energy expenditure while performing gymnastic-like motion in spacelab during spaceflight: case study. *Appl Physiol Nutr Metab 31*, 5 (Oct 2006), 631–635.
19. Khoshelham, K., and Elberink, S. O. Accuracy and resolution of kinect depth data for indoor mapping applications. *Sensors 12*, 2 (2012), 1437–1454.
20. Kozey, S. L., Lyden, K., Howe, C. A., Staudenmayer, J. W., and Freedson, P. S. Accelerometer output and met values of common physical activities. *Med Sci Sports Exerc 42*, 9 (Sep 2010), 1776–1784.
21. Krohn, M. M., and Boisdair, D. Use of a stereo-video system to estimate the energy expenditure of free-swimming fish. *Canadian Journal of Fisheries and Aquatic Sciences 51*, 5 (1994), 1119–1127.
22. Leatherdale, S. T., Woodruff, S. J., and Manske, S. R. Energy expenditure while playing active and inactive video games. *Am J Health Behav 34*, 1 (2010), 31–36.
23. Leprêtre, P. M., Weissland, T., Paton, C., Jeanne, M., Delannaud, S., and Ahmaidi, S. Comparison of 2 portable respiratory gas analysers. *Int J Sports Med 33*, 9 (Sep 2012), 728–33.
24. Levine, J. A. Measurement of energy expenditure. *Public Health Nutr 8*, 7A (Oct 2005), 1123–1132.
25. Maddison, R., Mhurchu, C. N., Jull, A., Jiang, Y., Prapavessis, H., and Rodgers, A. Energy expended playing video console games: an opportunity to increase children's physical activity? *Pediatr Exerc Sci 19*, 3 (Aug 2007), 334–343.
26. Mueller, F. F., Gibbs, M. R., and Vetere, F. Taxonomy of exertion games. In *OZCHI '08: Proceedings of the 20th Australasian Conference on Computer-Human Interaction*, ACM (New York, NY, USA, 2008), 263–266.
27. Osgnach, C., Poser, S., Bernardini, R., Rinaldo, R., and di Prampero, P. E. Energy cost and metabolic power in elite soccer: a new match analysis approach. *Med Sci Sports Exerc 42*, 1 (Jan 2010), 170–178.
28. Papastergiou, M. Exploring the potential of computer and video games for health and physical education: A literature review. *Computers & Education 53*, 3 (2009), 603 – 622.
29. Penko, A. L., and Barkley, J. E. Motivation and physiologic responses of playing a physically interactive video game relative to a sedentary alternative in children. *Ann Behav Med 39*, 2 (May 2010), 162–171.
30. Rosenberger, M. E., Skrinar, G., Haskell, W. L., Intille, S. S., and Tapia, E. M. Multiple wireless accelerometers and heart rate accurately predict energy expenditure during level walking. *Medicine and Science in Sports and Exercise, 40*, 5 (2008), S62–S63.
31. Rothney, M. P., Neumann, M., Béziat, A., and Chen, K. Y. An artificial neural network model of energy expenditure using nonintegrated acceleration signals. *J Appl Physiol 103*, 4 (Oct 2007), 1419–1427.
32. Schmidt, W. D., Biwer, C. J., and Kalscheuer, L. K. Effects of long versus short bout exercise on fitness and weight loss in overweight females. *J Am Coll Nutr 20*, 5 (Oct 2001), 494–501.
33. Shephard, R. J. A critical examination of the douglas bag technique. *J Physiol 127*, 3 (Mar 1955), 515–24.
34. Shotton, J., Fitzgibbon, A., Cook, M., Sharp, T., Finocchio, M., Moore, R., Kipman, A., and Blake, A. Real-Time Human Pose Recognition in Parts from Single Depth Images (June 2011).
35. Smola, A. J., and Schölkopf, B. A tutorial on support vector regression. *Statistics and Computing 14*, 3 (Aug. 2004), 199–222.
36. Su, S., Wang, L., Celler, B., Ambikairajah, E., and Savkin, A. Estimation of walking energy expenditure by using support vector regression. *Conf Proc IEEE Eng Med Biol Soc 4* (2005), 3526–3535.
37. Swartz, A. M., Strath, S. J., Bassett, Jr, D. R., O'Brien, W. L., King, G. A., and Ainsworth, B. E. Estimation of energy expenditure using csa accelerometers at hip and wrist sites. *Med Sci Sports Exerc 32*, 9 Suppl (Sep 2000), 450–456.
38. Tan, B., Aziz, A. R., Chua, K., and Teh, K. C. Aerobic demands of the dance simulation game. *Int J Sports Med 23*, 2 (Feb 2002), 125–134.
39. Vathsangam, H., Emken, A., Schroeder, E. T., Spruijt-Metz, D., and Sukhatme, G. S. Determining energy expenditure from treadmill walking using hip-worn inertial sensors: an experimental study. *IEEE Trans Biomed Eng 58*, 10 (Oct 2011), 2804–2819.
40. Xia, L., Chen, C.-C., and Aggrawal, J. K. View invariant human action recognition using histograms of 3d joints. In *The 2nd International Workshop on Human Activity Understanding from 3D Data (HAU3D) in conjunction with IEEE CVPR 2012,* (Providence, RI, 2012).