Observation of Selected Human Physiological Parameters Using Computer Vision

Gregor Koporec, Rok Mandeljc, Vildana Sulič Kenk, Janez Perš
University of Ljubljana, Faculty of Electrical Engineering
Tržaška cesta 25, 1000 Ljubljana
`gregor.koporec@gmail.com`
`janez.pers@fe.uni-lj.si`

Goran Vučković, Radoje Milić
University of Ljubljana, Faculty of Sport
Gortanova 22, 1000 Ljubljana

Abstract. *Measurement of physiological parameters is an important tool in sport science and medicine, especially when trying to estimate physical activity or energy consumption. However, most approaches still rely on sensors or markers, placed directly on the body. In this paper we explored the* observability *of selected physiological parameters using a fully contactless, fully automatic method, that relies on computer vision algorithms. We rely on the estimation of the optical flow to calculate Histograms of Oriented Optical Flow (HOOF) descriptors, which we subsequently augment with the histogram of flow magnitudes. The descriptors are fed into regression model, which allows us to estimate energy consumption and heart rate. Furthermore, with only minor modifications, mainly related to the scale of observation we are able to detect breathing of a sleeping person, confirming the generality of our approach. Our method has been tested both in lab environment and in realistic conditions, and the results confirm that these physiological parameters are indeed observable with purely visual contact and inexpensive equipment.*

1. Introduction

Measurement of physical activity is important not only for the determination of human health [23] but is also valuable in professional sport training and sports medicine. By measuring physical intensity we can predict the energy expenditures in individual sports [3, 16] and determine overloads, which can lead to muscle fatigue [18].

Due to the complexity, it is almost impossible to accurately measure the intensity of physical activity [23, 20]. The methods that are used for measurements are divided into direct and indirect calorimetry [2] and non calorimetric techniques such as heart rate monitors, pedometers and accelerometers [14]. The latter are widely used in research because they are easy to use [20], but their accuracy is questionable [23]. They are also impractical, since they restrict the movement of athletes and thus indirectly affect the result.

Because of these limitations, researchers have developed non-contact methods of intensity estimation based on the analysis of video [3, 16, 20, 17, 15]. Some of these methods use metabolic models, but none of them is using the motion field, derived from video.

We can reasonably expect that the full motion field gives the most accurate description of the physical activity. Our long-term goal is a set of methods and hardware devices to measure various physiological parameters from visual observations, physical activity being only one of them. However, our first task, which is at the core of this paper, is the analysis of *observability* of physiological parameters. We denote the parameters *observable* if there is non-zero (positive) correlation between our estimation of a parameter, based on motion features obtained from video and the values, obtained by other methods that are considered to be reliable for measurement for a particular parameter.

In Section 2 we present the methods, developed for non-contact measurement of physical activity. Next, we describe our proposed framework, together with its theoretical basis. The experiments are described in Section 4. At the end of the paper we present the experimental results and the future work.

2. Related work

Studies of contactless measurement of activity are rare, since most researchers are using contact measurement devices, for example [11]. In this work they were estimating energy consumption with regression models trained on multiple features from wearable sensors.

Nevertheless, some researchers tried to assess the energy consumption with contactless movement sensors, such as the Kinect [15]. Contactless sensors were also addressed in [20, 17], but the intensity was determined by the subjective scale.

In work [16], the authors proved that we can estimate energy difficulty of soccer with video analysis in conjunction with the metabolic model. They processed video frames and followed the player throughout the game. By tracking player movement they classified it in each category and calculated energy consumption based on physiological characteristics. However, this approach is possible only if majority of physical activity consists of cyclic motion (running).

A similar approach to measuring activities is presented in [3]. Authors determined aerobic power consumption for essential activities carried out in tennis and thus built a mathematical model of the metabolic process. They were able to determine the profile of activities on a video, and thus indirectly evaluate energy consumption.

The problem in these approaches are the limits of metabolic models – they are limited to a certain type of activity or sport. Additionally, the studies used offline analysis of videos, aimed at determining the total energy consumption during the game rather than real-time evaluation of this important physiological parameter. Finally, human annotation is expensive and slow. The automatic activity detection may alleviate this, but at a cost of introducing possibly significant sources of additional error and uncertainty.

3. Proposed method

The video camera is completely contactless device and therefore provides totally non-intrusive, optical way of measurement. Additionally, advances in technology now allow acquisition of high-definition videos, at high frame rates. Finally, optics allow extreme flexibility and observation from very large distance, unlike any other kind of sensor.

3.1. Movement model and features

To obtain detailed measurements of motion, we calculate the motion of each pixel – the *optical flow*. Under specific conditions (stable illumination) it represents a good approximation for the movement of the object's surface. Unfortunately, traditional methods of optical flow are sensitive to noise, discontinuity and changes in brightness of the object [4]. Additionally, there is a problem of correct estimation of the amplitude of movement due to parallax phenomenon [22]—objects that are further away from the camera exhibit lower flow amplitude for the same physical motion.

We decided that we will represent the movement from video recordings with histograms of oriented optical flow (HOOF) [7] to improve robustness. Reasoning behind that choice is illustrated by Figure 1 – HOOF features measure amount of motion in various directions and therefore they should accurately represent the motion of an athlete on a treadmill, as follows.
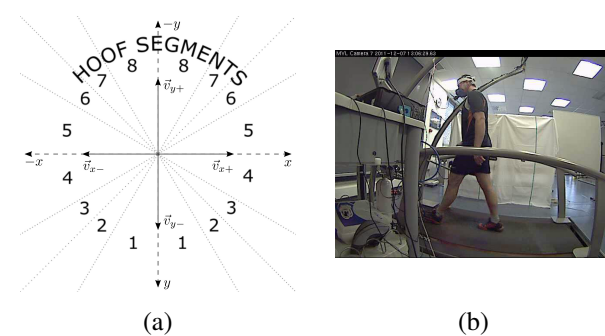

(a)   (b)

Figure 1: (a) Components of an athlete motion in world coordinates and illustration of their mappings onto HOOF features. For details, see text. (b) 150th frame of the first test from first series of videos. The treadmill speed in this case was $6\,\text{km}\,\text{h}^{-1}$.

Ideally, the movement of the particle with mass m can be decomposed into the components of velocity in the vertical and horizontal direction. Movement in the direction of $\vec{v}_{y+}$ represents the largest effort, because a muscle must defy the force of gravity to perform work. Conversely, it can move in the direction of $\vec{v}_{y-}$ relatively effortless. Movement in directions $\vec{v}_{x+}$ and $\vec{v}_{x-}$ is somewhere in between by the

invested energy. We can not accurately assess the work belonging to the movement in any direction, but we can *learn* it using HOOF features.

Following results in the realistic environment (a squash court), we decided to augment HOOF descriptor with the histogram of the absolute flow amplitudes, which significantly improved observed correlation between reference measurement and predicted values. In our implementation, we use $N_{HOOF} = 60$ segment HOOF feature vector, roughly representing 60 directions. This was later augmented with $N_{ampl} = 60$ amplitude bins, roughly corresponding to flow amplitudes in 0.5-60 pixel range, yielding 120-dimensional descriptor per each person and frame. Amplitude cut-off at 0.5 pixel was employed to get rid of noise in absence of motion.

3.2. Optical flow algorithm

Optical flow is calculated in full frames, but can be, depending on the situation, cropped with appropriate bounding box. For computation of the optical flow we have chosen Farnebäck's dense flow algorithm [9]. The main reason for this choice is the availability of its implementation in the OpenCV library.

3.3. Models

Based on the nature of the selected features we have chosen to employ models, based on support vector machines (SVM) [5]. Thus, despite the large number of features, we ensured a relatively short learning time and a degree of robustness. Based on recommendations from [13] we used a non-linear kernel.

4. Experiments

We present multiple experiments in varied environments, addressing measurement of instantaneous energy consumption, heart rate (considered the proxy for energy consumption) and breathing. In all cases, physiological parameters $eem(t)$, $hr(t)$, $breathing(t)$ were predicted from *one* optical flow image via feature vector at the time $\mathbf{X}(t - lag)$, where lag is the hypothesized delay between the moment certain activity is visible, and the moment it gets expressed in the physiological parameter. No other temporal modeling was used, except for final smoothing of predictions. This resulted in very simple, real-time model, which can be extended with

temporal modeling, should the need arise.

4.1. Treadmill experiment

The first set of experiments was performed in physiological laboratory, with subject running on a treadmill in the presence of the operator—a doctor, who determined the intensity and duration of workload. Heart rate and energy expenditure were measured for an athlete (age: 26 years, height: $177\,\mathrm{cm}$, weight: $79.1\,\mathrm{kg}$, VO_2max: $3705\,\mathrm{ml/\,min}$). Energy expenditure was measured using indirect calorimetry with Cosmed CPET Metabolic Cart. System allows breath-by-breath measurement [2]. We used Hans Rudolph face mask with prescribed minimal VD (dead space).

4.1.1 Data acquisition

We filmed the treadmill from the two different angles: the side-view and the back-view. The slope of the treadmill was from $1.5\,\%$ to $2\,\%$. We filmed in 480×640 resolution with a $30\,\mathrm{fps}$ speed. An example of a recording is shown in Figure 1(b).

4.1.2 Procedure

We have made two series of tests with 20 minutes between them. Physiological parameters were sampled every $5\,\mathrm{s}$. In the first series we made 8 tests, where every test lasted for 2 minutes. The treadmill's speed was increased by $1\,\mathrm{km\,h^{-1}}$ every test. First test had a speed of $6\,\mathrm{km\,h^{-1}}$ and the last a speed of $13\,\mathrm{km\,h^{-1}}$. In the second series we made 3 tests. Every test lasted for 5 minutes. Treadmill's speeds were $7\,\mathrm{km\,h^{-1}}$, $10\,\mathrm{km\,h^{-1}}$ and $13\,\mathrm{km\,h^{-1}}$. The first set was used for the acquisition of samples for learning, and the other for testing.

4.1.3 Processing

We then calculated the optical flow [9] with the help of tracking algorithm described in 4.2. For optical flow we used the following parameters: pyramid scale 0.5, number of pyramid layers 3, averaging window size 15, number of iterations at each pyramid level 3, size of the pixel neighborhood 5 and standard deviation of the Gaussian 1.2. An example of the obtained optical flow is shown in Figure 2.

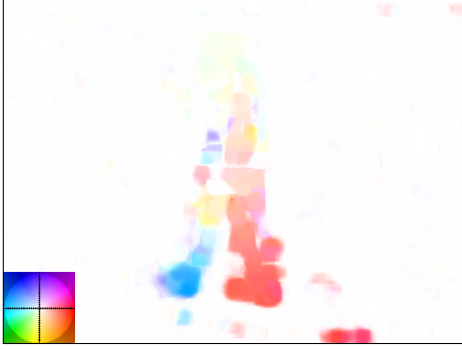HOOF features were calculated according to the method described in [7].

Figure 2: The optical flow for 150th frame of the first test from the first series of shots with color coding legend in the bottom left corner. We are using a standard color coding based on [1]. The maximum amplitude of the optical flow in this figure is $17\,\text{px}$.

The models have been generated with support vector regression ϵ-SVR in LIBSVM library, which is more specifically described in [5]. We used RBF kernel that takes the form (1). Kernel and regression parameters were optimized with grid search approach described in [13]. We needed to determine regression penalty parameter $C > 0$, loss function parameter $\epsilon > 0$, and kernel coefficient γ.

$$K(\mathbf{x_i}, \mathbf{x_j}) = e^{-\gamma \left\| \mathbf{x_i} - \mathbf{x_j} \right\|^2} \tag{1}$$

We have built 8 models, divided into two categories: *hr* models, which predict heart rate and *eem* models, which provide the energy expenditure in $\text{kcal}/\min$.

The models are further divided according to the type of input data. We used a side-view camera (abbreviation *sv*), and the back-view camera (abbreviation *bv*). We extended our experiment by incorporating lag between measurements and reference values. With models, marked as *lag*, we checked the proposed time delay between excitation and physiological response.

In experiments with *mixed* abbreviations, we built the model on the data from the one view, and tested it on the another view.

Additional *mixed* model experiment was generated with data from both cameras, side-view and back-view. Recordings from both cameras were concatenated and cropped.

4.2. Object tracking

In many sports, there are a number of players participating and therefore they are all visible in each video frame. Necessary component of such system will be a tracking functionality, therefore we ran a tracker on treadmill video to check how the method performs if the position of the player is non-stationary and obtained by the tracking algorithm. Results which included tracking step have *tr* abbreviation.

For object tracking we used KCF tracking framework implemented in OpenCV because it gave us best results. The tracking method is an implementation of [12] which is extended to KCF with color-names features. Extension is based on [8].

Default parameters for tracking were: Gaussian kernel bandwidth 0.2, linear interpolation factor for adaptation 0.075, regularization 0.01, max patch size 6400, spatial bandwidth 0.0625, resize features activated to improve the processing speed, training coefficients splitted into two matrices, wrapping around the kernel values not activated, non-compressed descriptors in gray, compressed descriptors in color-names, the PCA method to compress the features activated, compressed size 2 and compression learning rate 0.15.

With KCF tracking framework tracked objects were defined with bounding box. The region of interest, where bounding box was calculated, was set on the first frame of every recording. Bounding box was used to crop the region of interest from optical flow image of particular frame and calculated HOOF features on it. If tracker couldn't find an object—it disappeared from our view or there were technical difficulties to calculate correspondences—bounding box didn't exist and all histogram bins were therefore zero.

Finally, cameras may shake, if held manually. We simulated this scenario by artificially introducing random small displacements and rotation into the video. Every frame was transformed with random Euclidean transformation, where translation was limited to $4\,\%$ of frame size and rotation to $0.13\,\text{rad}$. Random transformations were smoothed with Kalman filter (2), where the variance of process noise was 2, the variance of model noise 1024 and variance for posteriori error covariance 2. The tracking algorithm was run *after* this motion noise was added, and these results are denoted by *sh* abbreviation.

Comparison between the measured physiological parameters ($5\,\text{s}$) sampling, and our predictions at $30\,\text{fps}$ required interpolation of physiological param-

eters, which was performed in Matlab.

4.3. Examining the lag in physiological response

We explored the problem of the lag in physiological response as well. Based on Figure 3, we found out that between the change in speed of the treadmill and change of the selected physiological parameter there is a delay. This is perfectly acceptable due to the nature of physiological processes. We found out that offset for heart rate is amounted to $15\,\text{s}$ and for energy expenditure to $55\,\text{s}$. Offset was included in models with the *lag* abbreviation.

4.4. Denoising the results

Because of the noisy output of models, we had to filter them with the Kalman filter [10]. It is represented by the equation in state space, where state is the state of speed v and acceleration a with unknown input parameters speed v_n and acceleration a_n. The initial velocity and acceleration were 0. Variance of process noise for all models was 0.04. Variance of model noise was 456.13. Due to the unknown initial values, we used variance 456.13 for the posteriori error covariance.

$$\begin{bmatrix} v(k+1) \\ a(k+1) \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} v(k) \\ a(k) \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} \begin{bmatrix} v_n(k) \\ a_n(k) \end{bmatrix} \quad (2)$$
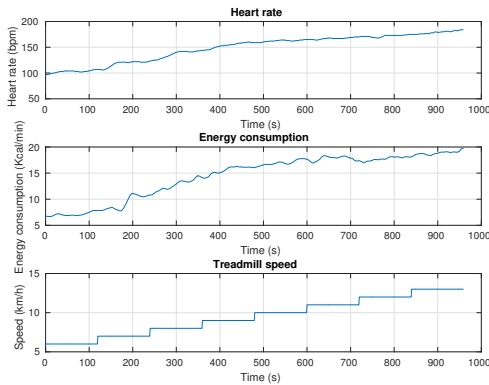

Figure 3: The figure shows the delay of physiological parameters response based on the treadmill speed.

4.5. Real-world squash experiment

The model squash match, consisting of only one set was filmed in 1920×1080 resolution with RaspberryPi and RaspiCam as a recording device. The heart rate was measured for both players using wearable sensors. First player (age: 45 years, height: 176 cm, weight: 68 kg, gender: male, max heart rate: 179 bpm, resting heart rate: 45 bpm) was used for training. Second player (age: 17 years, height: 178 cm, weight: 66 kg, gender: male, max heart rate: 203 bpm, resting heart rate: 50 bpm) was used for testing the model.

To obtain player bounding boxes, tracking [8] was employed, however the tracker was re-set once each 3 seconds by human operator to guarantee reasonable tracking results. We had to scale our frames to 25 % specifically for tracking, and remap the result to the original resolution later.

Poor initial performance with plain HOOF descriptors in a squash game prompted an extension of HOOF descriptor with amplitude histogram. This necessitated additional scaling step before building SVM model, where all features were scaled to the range $[-1, 1]$. Additionally, measured heart rate was first filtered with the Gaussian kernel of size 6 and variance 16 to prevent training on overly noisy data. It was then *individualized* to each player by calculating energy expenditure based on basic equation from [6]. Predicted results from model were then converted back to heart rate of the *other player* using the same equation. This allowed us to train the model on one player, and test it on another.

Kalman filter was not used for squash experiments. Because we used Gaussian kernel for filtering in data preprocessing, it was also used for filtering model output. The size of kernel was 6 samples and variance was 16.

4.6. Breathing detection

To show the generality of the proposed concept, we tested it on a loosely related problem of breathing detection. Different from sport applications, the use cases for such applications would be mainly in medicine, care for the elderly, or surveillance. The concept of optical measurement allows us to perform such measurements from great distance, as long as optical system is able to provide us with the stable image.

There are already many vision-based patient monitoring applications [19], one of which is also sleep apnea monitoring. As of [19] there are two main approaches to monitoring this disorder. One of these is tracking movement of chest region. However, our primary motivation was to test our proposed approach with minimum modifications on a different problem.

4.6.1 Method

For this purpose we recorded a video of a male subject, age 42, with history of diagnosed sleep apnea, when sleeping (recording started at 4:45 in the morning and lasted about 30 minutes, part of which was used). The illumination was provided by 60W near-infrared (NIR) LED illuminator, and recording was done again with RaspberryPi and RaspiCam (NIR version, without the NIR blocking filter). Filming was done in resolution at 25 fps which were reduced to 10 fps in video pre-processing to increase signal to noise ratio in calculated optical flow (breathing is a slow process). For the recording M12 lens with focal length of 1.8 mm was used (wide angle). Recording apparatus was approximately 2 meters from the observed subject.

4.6.2 Ground truth

To obtain reference values for breathing detection, sound was recorded as well using the audio module for RaspberryPi, with the microphone placed at close distance to the subject. Sound was synchronized to the video, and processed to obtain breathing detections based on high sound amplitude. By manual examination, it was established that the detections corresponded to the actual breathing, as heard on the sound track. Detections were subsampled to 10 samples per second, to coincide with the video frame rate.

4.6.3 Processing

To detect breathing, we observed a section of the subject's back (he was lying face down). That section, measured 384×512 pixels and covered approximately 2/3 of the subject's back. This was the only part of the image that was involved in any computation.

Two sections of video in duration of 5 minutes each were selected for training and testing, respectively. The training and testing were done using C-SVC classifier and RBF kernel with parameter optimization. To determine the performance, we formulated the problem as a binary classification problem with classes "no breathing" and "breathing".

5. Results

All energy consumption and heart rate models were validated on previously described test samples.

For comparison between the different models we have chosen validation measures: correlation coefficient (CORR), relative absolute error (RAE) and root relative square error (RRSE) [21]. The higher the value of the CORR the better, with RAE and RRSE is other way around.

Models were also evaluated with cross testing. This testing was done only by the type of input data—side-view or back-view. *sv* models, that were made with learning samples from side-view camera were first tested with testing samples from side-view camera and then with back-view camera. Hereafter tests with input data from side-view camera are marked with *sv* in brackets and tests with input data from back-view camera are marked with *bv* in brackets.

5.1. Treadmill results

As can be seen in the Table 1, we get relatively poor results in the prediction of heart rate.

Model	CORR	RAE (%)	RRSE (%)
hr-sv(sv)	0.90	75.42	76.66
hr-sv(bv)	−0.66	104.61	110.17
hr-sv-lag(sv)	**0.93**	**74.37**	**75.51**
hr-sv-lag(bv)	−0.87	138.60	136.62
hr-bv(sv)	0.83	311.95	295.66
hr-bv(bv)	0.88	81.07	79.27
hr-bv-lag(sv)	0.49	84.71	89.40
hr-bv-lag(bv)	0.91	79.15	76.93
eem-sv(sv)	0.87	47.08	49.75
eem-sv(bv)	−0.75	109.27	117.62
eem-sv-lag(sv)	**0.92**	**38.15**	**40.18**
eem-sv-lag(bv)	−0.74	121.86	121.88
eem-bv(sv)	0.61	94.92	95.92
eem-bv(bv)	0.85	44.51	56.24
eem-bv-lag(sv)	0.86	56.60	68.61
eem-bv-lag(bv)	0.90	45.51	49.50

Table 1: The results of the initial model evaluations with cross testing. For each model, we calculated the correlation coefficient (CORR), relative absolute error (RAE) and root relative square error (RRSE).

In terms of using different view angles, we get best results for side-view. We assume that this is due to the fact that with a back-view camera (without any cropping) we also recorded the movement of the operator, who is not visible in side-view recordings. Despite the fact that the HOOF features get rid of the noise of the optical flow, the movement of the operator is intensive enough to be able to influence the results. Worse results for back-view camera could also indicate that we get less descriptive features from it.

The results of the models, where we assume the delay between excitation and response are better, which indicates that the assumption is justified.

Considering cross testing we can see, that all models produce poorer results if we test them with data from different viewing angle.

The best results were obtained in the prediction of the energy expenditure (EEM). Output signals of the best results for the prediction of energy expenditure and heart rate are presented in Figure 4. The curves of the results have the same form because they are correlated physiological parameters.
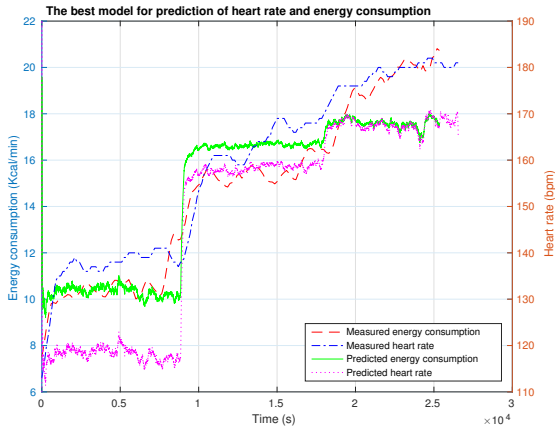

Figure 4: The best results for prediction of energy expenditure and heart rate when validating the models. The figure shows the output of models eem-sv-lag(sv) and hr-sv-lag(sv) and the actual value of energy expenditure and heart rate.

5.2. Mixed-view experiments

As in 5.1, when comparing models with different physiological parameters in Table 2, heart rate models produce worse results. Lag models are better than normal models and best result is still produced by lagged model, which predicts energy expenditure.

The main difference in mixed models can be seen, when comparing cross tests. If we compare results from Table 1 and 2, we can see that results, when testing models with data from different viewing angle as they were trained, are significantly better. This results indicate that better models could be trained with recordings from different viewing angle.

5.3. Treadmill with tracking

Results of models with enabled tracker are represented in Table 3. If we compare them with results of initial models in Table 1, the mean absolute difference of RRSE between them is $28\,\%$. We can assume

Model	CORR	RAE (%)	RRSE (%)
hr-mixed(sv)	0.89	67.18	68.17
hr-mixed(bv)	0.88	59.84	61.89
hr-mixed-lag(sv)	0.92	65.24	66.44
hr-mixed-lag(bv)	**0.91**	**57.75**	**60.31**
eem-mixed(sv)	0.85	45.90	53.89
eem-mixed(bv)	0.84	57.44	62.78
eem-mixed-lag(sv)	**0.90**	**44.19**	**46.09**
eem-mixed-lag(bv)	0.89	56.70	55.04

Table 2: The results of the mixed model evaluations with cross testing. For each model, we calculated the correlation coefficient (CORR), relative absolute error (RAE) and root relative square error (RRSE).

that this is due to the fact that tracker does not track selected object perfectly. In some cases it cannot find object, or detects wrong object. It can also track only part of the object. This anomalies can affect calculation of physiological parameters.

The mean absolute difference of RRSE between normal tracking models and models with shaking video is about $30\,\%$. Results are worse with videos that incorporate shaking (motion noise), but this is still acceptable, because the selected tracker can stabilize our video and improve results.

5.4. Squash match experiments

If we compare Table 1 and Table 4, we can see that result for squash model is not far behind one of the best results in initial models, despite the fact that we used different subjects for training and testing.

If we further explore our model from realistic data, we can see in Figure 5 that predicted working point is about $7\,\mathrm{bpm}$ lower. The reason could be that training data is extracted only from one subject. The second reason could be imperfections of used equation from [6]. Despite these errors a coarse prediction is still possible, even if we don't train the algorithm on the same subject.

5.5. Breathing experiment

For breathing detection, which was formulated as a classification problem, we used standard metrics for evaluation of two-class classification problems. With "breathing" considered the "true" value, and "not breathing" the "false" we get the following results: false positive rate, FPR = $13\,\%$, true positive rate, TPR = $87\,\%$, false negative rate, FNR = $26\,\%$, true negative rate, TNR = $74\,\%$.

Model	CORR	RAE (%)	RRSE (%)
hr-sv-tr(sv)	0.93	90.82	86.55
hr-sv-tr(bv)	−0.18	133.17	145.74
hr-sv-lag-tr(sv)	**0.96**	**91.57**	**86.72**
hr-sv-lag-tr(bv)	−0.11	108.99	124.40
hr-bv-tr(sv)	−0.55	132.25	146.66
hr-bv-tr(bv)	0.89	116.38	111.78
hr-bv-lag-tr(sv)	−0.62	131.09	140.59
hr-bv-lag-tr(bv)	0.91	118.69	113.24
eem-sv-tr(sv)	0.90	41.55	45.25
eem-sv-tr(bv)	−0.34	135.25	141.63
eem-sv-lag-tr(sv)	**0.94**	**31.66**	**37.05**
eem-sv-lag-tr(bv)	0.65	126.00	130.04
eem-bv-tr(sv)	−0.44	107.47	107.91
eem-bv-tr(bv)	0.91	53.21	52.92
eem-bv-lag-tr(sv)	−0.68	110.55	113.64
eem-bv-lag-tr(bv)	0.93	41.57	51.53
hr-sv-tr-sh(sv)	0.92	90.39	87.15
hr-sv-tr-sh(bv)	0.84	90.98	112.00
hr-sv-lag-tr-sh(sv)	**0.95**	**88.86**	**86.99**
hr-sv-lag-tr-sh(bv)	−0.10	111.46	118.79
hr-bv-tr-sh(sv)	0.83	286.16	268.48
hr-bv-tr-sh(bv)	0.87	113.11	111.15
hr-bv-lag-tr-sh(sv)	0.89	293.45	275.83
hr-bv-lag-tr-sh(bv)	0.87	114.98	113.90
eem-sv-tr-sh(sv)	0.90	50.18	59.92
eem-sv-tr-sh(bv)	0.89	119.57	128.17
eem-sv-lag-tr-sh(sv)	0.93	51.47	56.55
eem-sv-lag-tr-sh(bv)	−0.08	135.85	133.27
eem-bv-tr-sh(sv)	0.75	179.11	172.30
eem-bv-tr-sh(bv)	0.90	52.85	54.43
eem-bv-lag-tr-sh(sv)	0.91	175.29	171.63
eem-bv-lag-tr-sh(bv)	**0.94**	**50.02**	**48.93**

Table 3: The results of the tracker model evaluations with cross testing. For each model, we calculated the correlation coefficient (CORR), relative absolute error (RAE) and root relative square error (RRSE).

Model	CORR	RAE (%)	RRSE (%)
hr-bv-lag-tr-sq	0.45	68.65	54.24

Table 4: The results of the squash match evaluation. For model, we calculated the correlation coefficient (CORR), relative absolute error (RAE) and root relative square error (RRSE).

6. Discussion

In this paper we explored the *observability* of selected physiological parameters using a fully contactless and automatic method, that relies on computer vision algorithms. We combined HOOF features with SVM regression and classification to predict energy expenditure, heart rate, and breathing of the observed subject. We supplemented those fea-
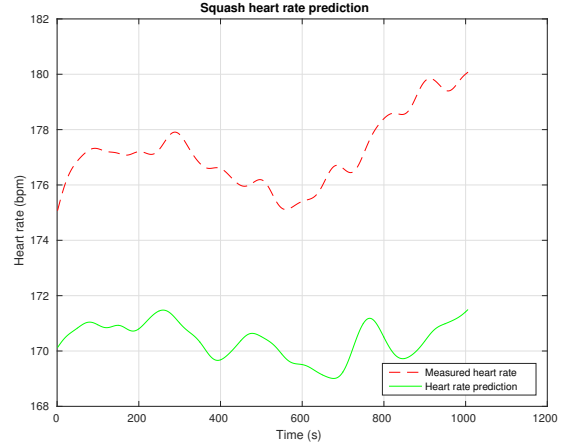

Figure 5: Result for prediction of heart rate when validating the models. The figure shows the output of model hr-bv-lag-tr-sq and the actual value of heart rate.

tures with a set of amplitude histograms to better represent amplitude of motion in our feature data.

Results show that the parameters are indeed observable, as we suspected. It is important to note that this was an *observability* study, which is the first step towards designing robust and accurate methods for measuring the parameters in question. Reason that this research is preliminary are as follows.

Methodology. We did not aim to find the best methodology to measure energy expenditure via computer vision. Based on the obtained results, we merely concluded that such measurements are possible. Furthermore, except for the smoothing, we do not incorporate any temporal dynamics into our model.

Groud truth. The reference methods that we used for comparison with our predictions suffer from numerous and documented problems. Heart rate is poor proxy for energy expenditure and is not used for that purpose in scientific research. It is widely used in training, since it is the only practical option for use in realistic sport environment, the other being bulky indirect calorimetry backpack with face mask.

Our future work will therefore consist of formalization and optimization of our augmented HOOF descriptor. We will explore SVM kernels that are better suited for histogram-type descriptors, and will introduce temporal dynamics into the model. Finally, we will focus on verification of our approach using the most accurate reference values we can obtain in the field of sport physiology and medicine.

References

[1] S. Baker, D. Scharstein, J. Lewis, S. Roth, M. J. Black, and R. Szeliski. A database and evaluation methodology for optical flow. *International Journal of Computer Vision*, 92(1):1–31, 2011. 4

[2] W. L. Beaver, K. Wasserman, B. J. Whipp, et al. On-line computer analysis and breath-by-breath graphical display of exercise function tests. *Journal of Applied Physiology*, 34(1):128–132, 1973. 1, 3

[3] F. Botton, C. Hautier, and J.-P. Eclache. Energy expenditure during tennis play: a preliminary video analysis and metabolic model approach. *The Journal of Strength & Conditioning Research*, 25(11):3022–3028, 2011. 1, 2

[4] T. Brox and J. Malik. Large displacement optical flow: descriptor matching in variational motion estimation. *IEEE transactions on pattern analysis and machine intelligence*, 33(3):500–513, 2011. 2

[5] C.-C. Chang and C.-J. Lin. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2:27:1–27:27, 2011. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm. 3, 4

[6] K. Charlot, J. Cornolo, R. Borne, J. V. Brugniaux, J.-P. Richalet, D. Chapelot, and A. Pichon. Improvement of energy expenditure prediction from heart rate during running. *Physiological measurement*, 35(2):253, 2014. 5, 7

[7] R. Chaudhry, A. Ravichandran, G. Hager, and R. Vidal. Histograms of oriented optical flow and binet-cauchy kernels on nonlinear dynamical systems for the recognition of human actions. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 1932–1939. IEEE, 2009. 2, 3

[8] M. Danelljan, F. Shahbaz Khan, M. Felsberg, and J. Van de Weijer. Adaptive color attributes for real-time visual tracking. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1090–1097, 2014. 4, 5

[9] G. Farnebäck. Two-frame motion estimation based on polynomial expansion. In *Scandinavian conference on Image analysis*, pages 363–370. Springer, 2003. 3

[10] D. A. Forsyth and J. Ponce. *Computer vision: a modern approach*. Prentice Hall Professional Technical Reference, 2002. 5

[11] H. Gjoreski, B. Kaluža, M. Gams, R. Milić, and M. Luštrek. Context-based ensemble method for human energy expenditure estimation. *Applied Soft Computing*, 37:960–970, 2015. 2

[12] J. F. Henriques, R. Caseiro, P. Martins, and J. Batista. Exploiting the circulant structure of tracking-by-detection with kernels. In *European conference on computer vision*, pages 702–715. Springer, 2012. 4

[13] C.-W. Hsu, C.-C. Chang, C.-J. Lin, et al. A practical guide to support vector classification. 2003. 3, 4

[14] J. A. Levine. Measurement of energy expenditure. *Public health nutrition*, 8(7a):1123–1132, 2005. 1

[15] D. Nathan, D. Q. Huynh, J. Rubenson, and M. Rosenberg. Estimating physical activity energy expenditure with the kinect sensor in an exergaming environment. *PloS one*, 10(5):e0127113, 2015. 1, 2

[16] C. Osgnach, S. Poser, R. Bernardini, R. Rinaldo, and P. E. Di Prampero. Energy cost and metabolic power in elite soccer: a new match analysis approach. *Med Sci Sports Exerc*, 42(1):170–178, 2010. 1, 2

[17] K. A. Peker and A. Divakaran. Framework for measurement of the intensity of motion activity of video segments. *Journal of Visual Communication and Image Representation*, 15(3):265–284, 2004. 1, 2

[18] T. Reilly. Energetics of high-intensity exercise (soccer) with particular reference to fatigue. *Journal of sports sciences*, 15(3):257–263, 1997. 1

[19] S. Sathyanarayana, R. K. Satzoda, S. Sathyanarayana, and S. Thambipillai. Vision-based patient monitoring: a comprehensive review of algorithms and technologies. *Journal of Ambient Intelligence and Humanized Computing*, pages 1–27, 2015. 5

[20] P. Silva, C. Santiago, L. Reis, A. Sousa, J. Mota, and G. Welk. Assessing physical activity intensity by video analysis. *Physiological measurement*, 36(5):1037, 2015. 1, 2

[21] I. H. Witten and E. Frank. *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann, 2005. 6

[22] L. Xu, Z. Dai, and J. Jia. Scale invariant optical flow. In *Computer Vision–ECCV 2012*, pages 385–399. Springer, 2012. 2

[23] K. Zhang, F. X. Pi-Sunyer, C. N. Boozer, et al. Improving energy expenditure estimation for physical activity. *Medicine and Science in Sports and Exercise*, 36(5):883–889, 2004. 1