Measuring energy expenditure in sports by thermal video analysis

Rikke Gade
Section of Media Technology
Aalborg University, Denmark
rg@create.aau.dk

Ryan Godsk Larsen
Department of Health Science and Technology
Aalborg University, Denmark
rl@hst.aau.dk

Thomas B. Moeslund
Section of Media Technology
Aalborg University, Denmark
tbm@create.aau.dk

Abstract

Estimation of human energy expenditure in sports and exercise contributes to performance analyses and tracking of physical activity levels. The focus of this work is to develop a video-based method for estimation of energy expenditure in athletes. We propose a method using thermal video analysis to automatically extract the cyclic motion pattern, in walking and running represented as steps, and analyse the frequency. Experiments are performed with one subject in two different tests, each at 5, 8, 10, and 12 km/h. The results of our proposed video-based method is compared to concurrent measurements of oxygen uptake. These initial experiments indicate a correlation between estimated step frequency and oxygen uptake. Based on the preliminary results we conclude that the proposed method has potential as a future non-invasive approach to estimate energy expenditure during sports.

1. Introduction

Measuring the performance of athletes in sports is of interest at many levels. In elite and professional sports athletes are regularly tested for their maximum performance and endurance level, and at amateur levels performance of both organised sports and daily exercises are measured and compared with team mates and friends. Self-tracking has become a trend in all aspects of life.

Due to the sedentary everyday life of most people in the modern society, both governments and individuals focus on the amount of physical activity performed regularly. Keeping track of ones performance might inform and inspire a healthy lifestyle.

In this paper we will focus on the estimation of energy expenditure, which quantifies the activity level of humans. Traditionally, in sports science, energy expenditure is measured directly by oxygen uptake [7]. This requires the participant to wear a mask connected to a stationary metabolic cart while performing tests on a treadmill or cycle ergometer, or in case of free movement the athlete must carry a mobile metabolic cart. As this equipment is both expensive and may interfere with the natural behaviour of the athlete, the method is primarily applied in science and when testing professional athletes. For daily activities and amateur sports cheaper methods with inexpensive devices are applied. These indirect methods for estimating energy expenditure include heart rate registration [11], accelerometry [5], and GPS [8], which have been validated against oxygen uptake. Particularly accelerometers and GPS in wearable devices or smartphones are often used by individuals wanting to track their own activities on a daily basis. However, these methods also require special equipment attached directly to athletes. Each technology has limitations; heart rate measurements may introduce inaccuracy because of emotional or environmental stress, and the training status can affect the association between oxygen uptake and heart rate [5, 11]. GPS can only be applied in outdoor environments, and studies show that energy expenditure can not be precisely estimated from this technology [8]. Accelerometers seem to be the best low-cost wearable technology for this purpose, but especially uniaxial accelerometers have limitations at high speeds [5].

The focus of this paper is on estimation of energy expenditure for sports activities with a new non-invasive camera-based method.

2. Related work

During recent years the focus on vision-based analysis of sports has increased significantly in both research and commercial systems [15]. Particularly, tracking the motion of players is an important step in the analysis of, e.g., physical performance and tactics [2, 17]. Going further into the analysis of human motion often involves research in estimation of body pose [13].

Even though the distance covered, found by tracking, or the recognised activity relates to energy expenditure of humans, only few papers compares vision-based measurements with energy expenditure directly. Tsou and Wu [16] exploited a Microsoft Kinect sensor to remotely track the movements of 10 joints and they estimate the energy expenditure during aerobics exercises. This approach requires reliable skeleton data, which currently only operates reliably at close distances, and when the subject is facing the camera. Kim et al. [10] presented a similar approach, using a Microsoft Kinect sensor, for assessing the energy expenditure of subjects playing exercise games.

Osgnach et al. [12] applied a semiautomatic visual tracking system to gather performance data from soccer matches. From players' speed, accelerations and decelerations energy expenditure was estimated. From 2D images captured by a cell phone camera Yang et al. [18] estimated the energy expenditure of indoor workouts (sit up, push up, jumping jack, and squat). This method is based on estimation of repetitions and intensity level of each activity.

Non-invasive video-based energy expenditure measurement has also been investigated for rehabilitation purposes and assistive monitoring for elderly people. In applications like these, activities can be assumed to have low intensity and slow speed, which causes the duration of activities to be more important than the momentary activity level. Edgcomb and Vahid [4] estimated energy expenditure in normal household activities from the sum of horizontal accelerations of a person captured by video. Tao et al. [14] used RGB-Depth sensors for estimating energy expenditure during daily living activities at home. First, the activity type was recognised, after which the energy expenditure was estimated.

A first step towards measuring energy expenditure from thermal video was described in [9]. The purpose of this paper was to investigate the correlation between energy expenditure measured by oxygen uptake and a video based motion measure. For a constrained scenario with horizontal walking and running on a treadmill, energy expenditure was estimated from optical flow, which proved to have a linear correlation with oxygen uptake. As this is a valid method when assuming a static treadmill setup, it has limitations for free movement.

The previous work published in this field shows that it is possible to estimate energy expenditure from video. However, previous methods applied on sports activities either require close distance 2D or depth images at a fixed angle, or manual interventions in a semiautomatic tracking system. In this work we will focus on general sports activities observed from a distance, e.g., in an indoor sports arena or at an outdoor field.

3. Approach

With a camera based method, our goal is to develop a non-invasive estimation method of energy expenditure for general sports activities. Inspired by the nature of accelerometer data, we aim to extract the dynamics of human motion, but by using video data we avoid the need of extra equipment attached to each individual.

In this work we focus on controlled motion patterns, in particular walking and running at constant velocity. With this approach we are able to directly compare our results with steady state measurements of oxygen uptake at each level of activity. As a direct continuation of the treadmill experiments published in [9] we start our experiments by considering motion in a straight line with a camera capturing a side-view. After that, we continue by examining running activities in a circular pattern, for which the viewing angle on the body will continuously change.

4. Methods

The first step to consider is the image acquisition technique. In this work we will use a thermal camera to avoid any privacy issues while testing in public sports arenas. However, the methods proposed will be generic with only minor adjustments to different image types. For applications where the identity of individuals is of interest, RGB video would be preferred.

A thermal camera captures infrared radiation in the mid- or long-wavelength infrared spectrum depending on sensor type (approx. 3-5 μm and 7-15 μm, respectively). This radiation, often called thermal radiation, is emitted from any object with a temperature above absolute zero, where intensity and dominating wavelengths depend on the temperature [6]. As such, pixel values represent temperature, in this paper visualised with white for the hottest pixels and black for the coldest pixels. An example of input image used is shown in figure 1.

4.1. Segmentation

Using a static camera view, background subtraction can be applied to detect the foreground. Assuming that the only moving objects in the scene are humans, the foreground will consist of people and possibly noise, which includes reflections of thermal radiation in the floor. After thresholding the difference image a binary image is obtained, which can still contain noise. An example of the resulting difference image from background subtraction, and the binary result after thresholding the difference image are shown in figure 2(b) and (c).

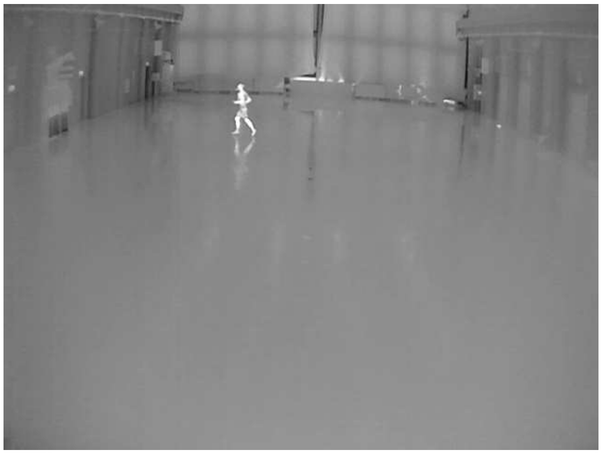


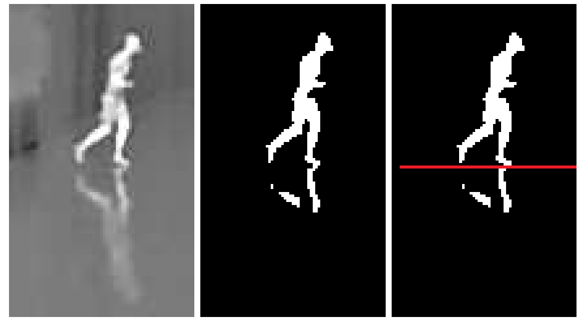
Figure 1: Example of thermal image from a sports arena.

(a) Thermal (b) Difference (c) Binary (d) Closed

Figure 2: Input image (a) and results of three segmentation steps: (b) background subtraction, (c) thresholding (threshold value 50), and (d) morphology (closing). All images are cropped.

Smaller noise objects can be filtered by size, however, there is also a risk that pixels from true body parts are detached from the main part when thresholding the image. They may then be interpreted as noise pixels. This is the case for one leg in figure 2(c). In order to close small gaps and holes, and aim for one connected white object for each person, we apply morphology closing on a binary image using a structuring element of disk shape with 5 pixels diameter. The result is shown in figure 2(d).

As it can be observed in figure 2, the reflection of thermal radiation from the body in the floor can be detected. In cases where the size is small and reflections are separated from the body, it can be filtered and left out from further processing. But in other cases the reflection might be connected to the body. In order to separate reflections from the body, the binary object can be analysed, and a decision on if and where to cut the object is made. From empirical studies of segmentation results, it is decided to cut the object if a row with less than five white pixels is found in the second lowest sixth of the bounding box area. The lowest part of the area is not used, due to the high risk of cutting a foot from a good segmentation. An illustration of this step is shown in figure 3.



(a) Thermal (b) Binary (c) Binary with cut

Figure 3: Illustration of how to separate body from reflection.

4.2. Estimation of energy expenditure

From the segmented body silhouettes, we aim to develop a method that captures the dynamics of human motion, independent of viewing angle and distance covered by the person. For this initial work we focus on the cyclic motion pattern of natural human motion like walking and running. We will show that it can be extracted by analysing the bounding box of the person's silhouette. Example images from one step cycle of a test person running at 10 km/h is provided in figure 4.

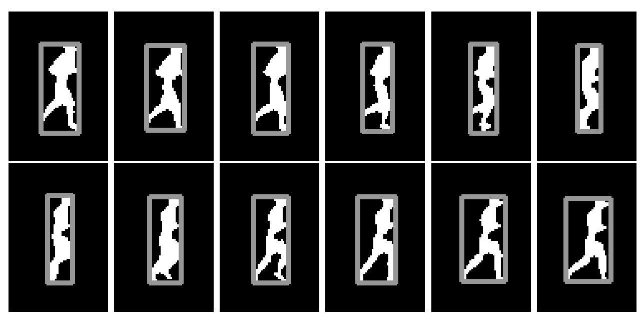


Figure 4: Sequence of frames during one step cycle with bounding boxes marked. All images are cropped.

As seen in figure 4 the bounding box primarily changes the width when observing a person from a side-view or partly side-view. However, from a front-view the height of the bounding box contributes to the cyclic pattern. Furthermore, as the distance between camera and person changes, the pixel dimensions of a person in the image changes. Therefore, the ratio between height and width of the bounding box may better represent the motion pattern independently of scaling. An example of bounding box ratio measured during three steps at 10 km/h (captured at a framerate of 30 fps) is presented in figure 5.

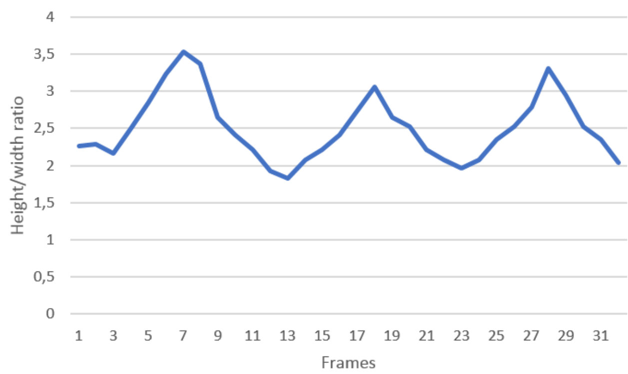


Figure 5: Example of bounding box height/width ratio during three steps when running at 10 km/h, captured at 30 fps.

Maximum points of the ratio graph over time represents steps. These are extracted from the slope between data points; a maximum is registered when the slope changes from positive to negative sign. The frequency of maximums corresponds to the step frequency.

5. Experiments

The developed video based method is evaluated against oxygen uptake, which is the direct quantification of energy expenditure, as discussed in section 1. The experiments described in this paper are conducted with one participant, and can therefore not be considered a full validation of the method, but should be seen as a proof-of-concept which leads to further research and evaluations.

The experiment is divided into two tests with different running protocols. During the first test the participant runs along a straight line of 20 metres and turns at each end, as illustrated in figure 7. A frame from this test is shown in figure 1. The participant is told to keep a constant velocity, which is controlled by a sound signal when the participant should reach each end point.

During the second test the participant runs along a circle of 20 metres diameter as illustrated in figure 8. The velocity is controlled by a sound signal which should be reached at fix points placed at each eighth of the circle perimeter. Frames exemplifying different views from the circle pattern is presented in figure 6. For both tests, the camera is placed in a height of 3 metres.

Each test is performed at four different velocities: 5 km/h (walking), 8, 10, and 12 km/h (running). The participant runs a 4-minute interval at each velocity, with rest time between each interval. 4 minutes of constant velocity is chosen to ensure that a steady state oxygen uptake is reached. Only the last minute of each interval is used for processing of both oxygen and video data.

Video is captured at 30 fps with a thermal camera of type

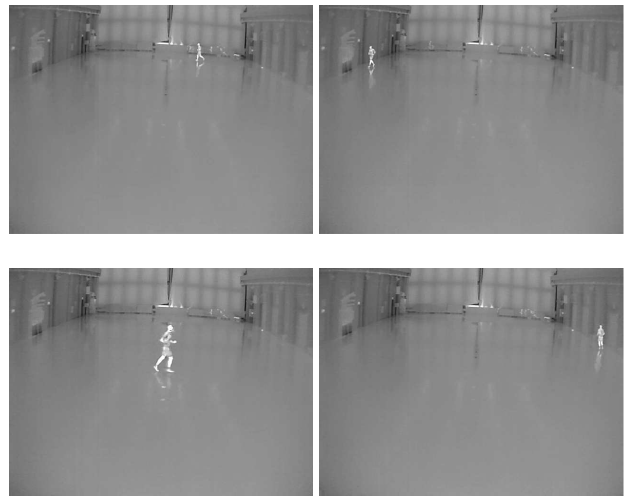


Figure 6: Frame examples from test 2, running along a circle with a diameter of 20 metres.

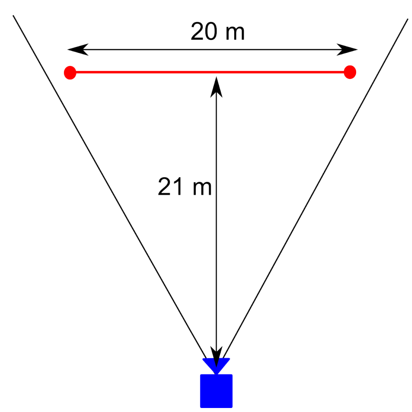


Figure 7: Sketch of test setup for test 1. Participant runs along the red line and turns at each end point.

AXIS Q1922 with a 10mm lens, 57 degrees field-of-view, 640 $\times$ 480 pixels [1]. Oxygen uptake is measured with a mobile Cardiopulmonary Exercise Testing (CPET) system, Carefusion - Oxycon mobile by Jaeger [3]. The test participant is an endurance trained male subject, with a weight of 69 kg.

5.1. Results

For each 1-minute video sequence, the bounding box ratio of the detected person is extracted and saved. The results are presented in figures 11 and 12 for line and circle patterns, respectively. For the first test, running in a straight line, the camera observes the athlete from a side-view, which provides a relatively stable ratio pattern. This is, however, interrupted every time the athlete turns, which can be observed 4 times at 5 km/h, 7 times at 8 km/h, 9

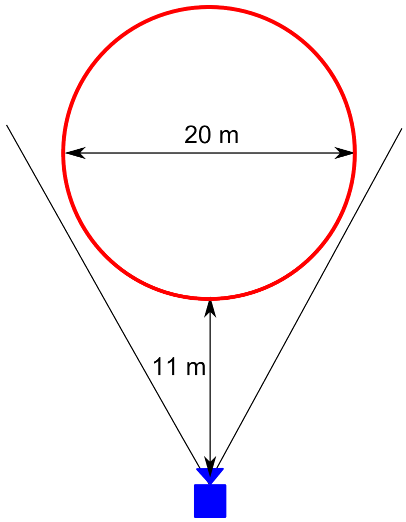


Figure 8: Sketch of test setup for test 2. Participant runs along the red circle.

times at 10 km/h, and 10 times at 12 km/h. When running in a circle the magnitude variations in the ratio measurements depend on the viewing angle. When observing the athlete directly from the front or from the back, only small variations are observed in the ratio.

From these ratio graphs local maximums are extracted as described in section 4.2. Each maximum can be interpreted as a step or cyclic repetition of motion. Table 1 summarises the number of detected maximums and the gross oxygen uptake measured at the same sequence. Figures 9 and 10 plot these results.

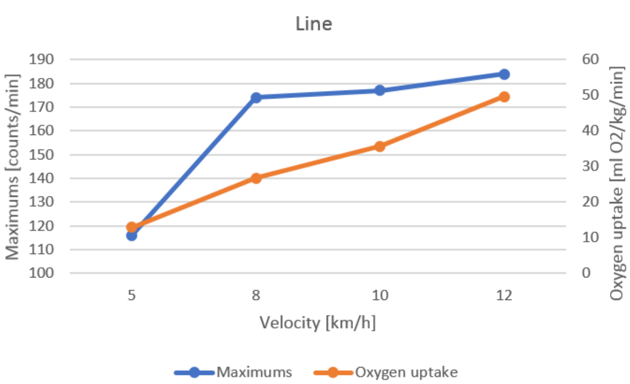


Figure 9: Results of test 1.

The comparisons between maximum counts and oxygen uptake, plotted against velocity, indicates a correlation. In test 1 the number of maximums shows a large difference between 5 km/h and 8 km/h, which is not observed from oxygen data. However, for test 2, when the participant is running continuously in a circle, a linear correlation is observed. To conclude whether these correlation patterns are generic we will need a test with a larger number of participants.

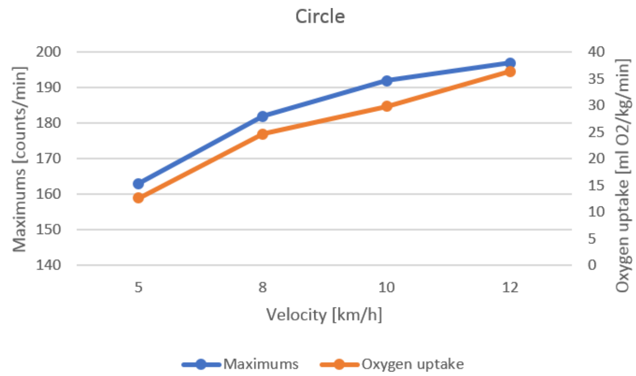


Figure 10: Results of test 2.

6. Discussion

In this work we have investigated the estimation of energy expenditure from visual analysis of thermal video. Tested at two different running protocols, each at four different velocities, we compared the extracted repetition numbers with oxygen uptake.

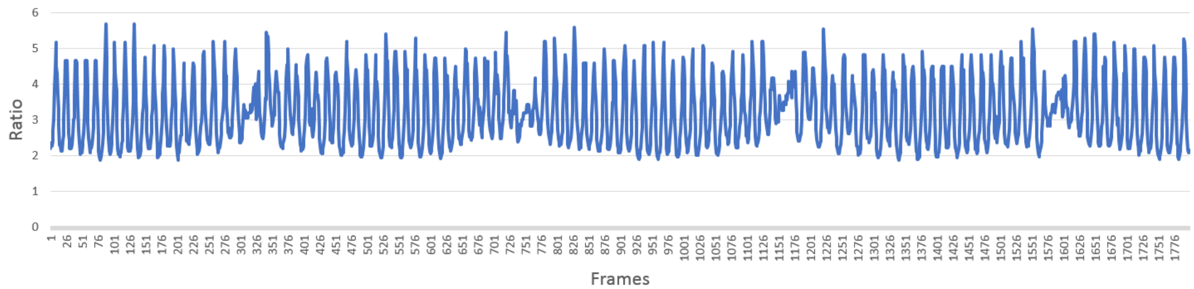
Considering oxygen uptake as the ground truth quantification of energy expenditure, we observe that the two different test scenarios have different energy costs. At a velocity of 5 km/h the difference in oxygen uptake between line and circle patterns is insignificant, but at higher velocities the line pattern requires more energy than the circle pattern. This can be related to the 180 degrees turn that happens for every 20 metres, which forces the athlete to decelerate and accelerate. At the circle pattern the velocity is kept constant. These abrupt changes when turning do also cause noise in the visual analysis, which can be seen in figure 11.

We measure the changes in ratio of the bounding box, which is interpreted as the step frequency of running, but could also be repetitions in still-standing exercises with similar repetitive motion patterns. For running, two factors influence the speed, hence the energy expenditure; step frequency and stride length. For future work it could therefore be interesting to investigate whether relative stride length or similar intensity level of activities can improve the estimation of energy estimation.

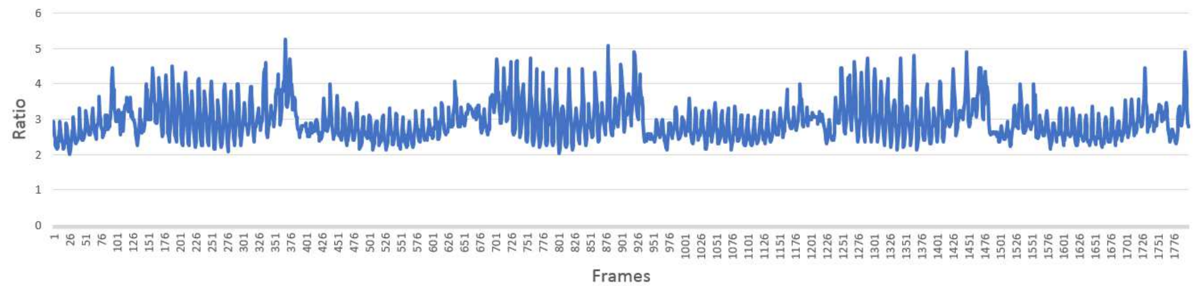
The initial experiments presented in this paper include only one participant due to the demanding test protocol when measuring oxygen uptake. However, we have proved that a relatively simple visual analysis can provide estimations of energy expenditure which are related to the measurements of oxygen consumption. It therefore serves as a proof-of-concept, which will be the starting point for further research and a thorough evaluation. In addition to more test

		5 km/h	8 km/h	10 km/h	12 km/h
Line	Gross oxygen uptake [ml O_2/kg/min]	13.0	26.8	35.7	49.6
	Maximums [counts/min]	116	174	177	184
Circle	Gross oxygen uptake [ml O_2/kg/min]	12.7	24.7	29.8	36.4
	Maximums [counts/min]	163	182	192	197

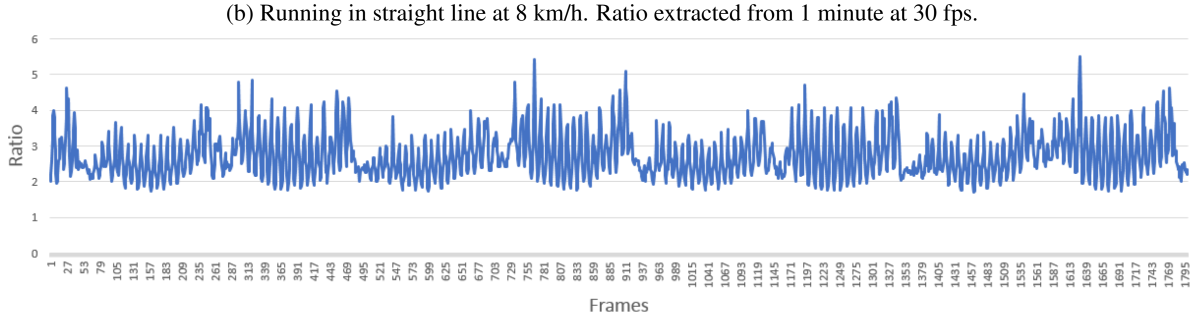
Table 1: Oxygen uptake and step counts from both tests.



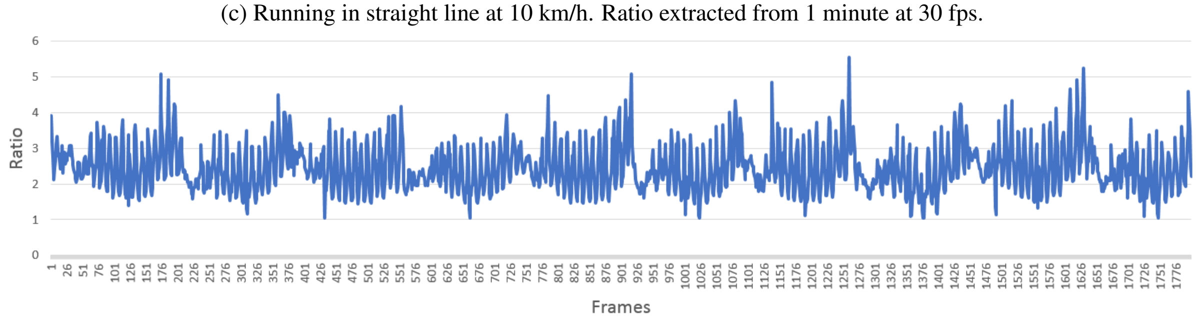
(a) Walking in straight line at 5 km/h. Ratio extracted from 1 minute at 30 fps.



(b) Running in straight line at 8 km/h. Ratio extracted from 1 minute at 30 fps.

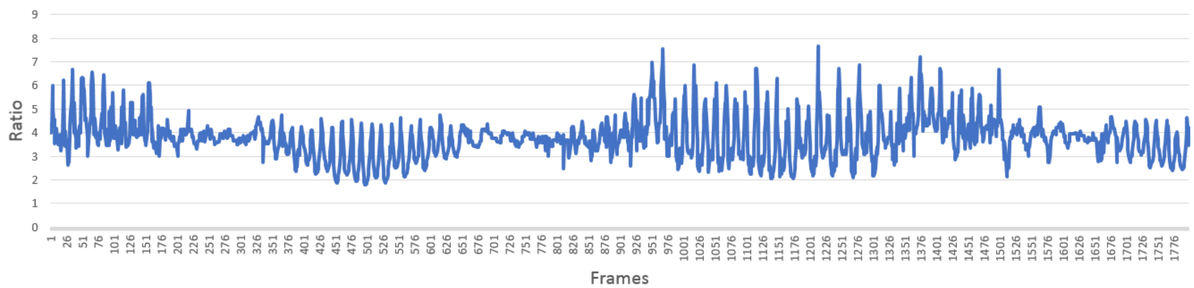


(c) Running in straight line at 10 km/h. Ratio extracted from 1 minute at 30 fps.

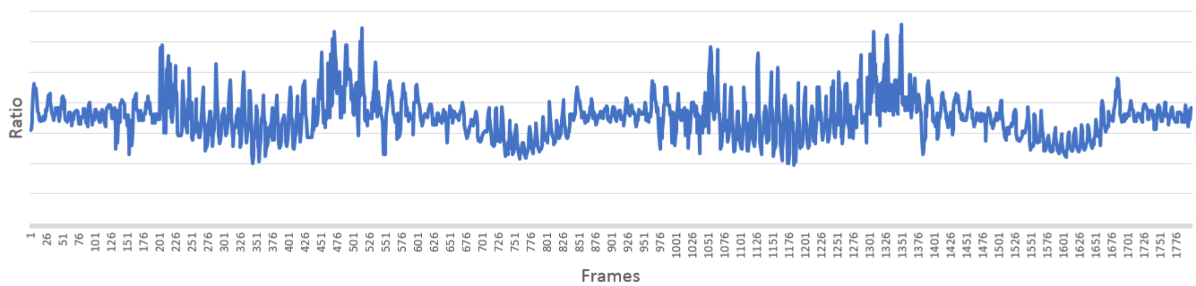


(d) Running in straight line at 12 km/h. Ratio extracted from 1 minute at 30 fps.

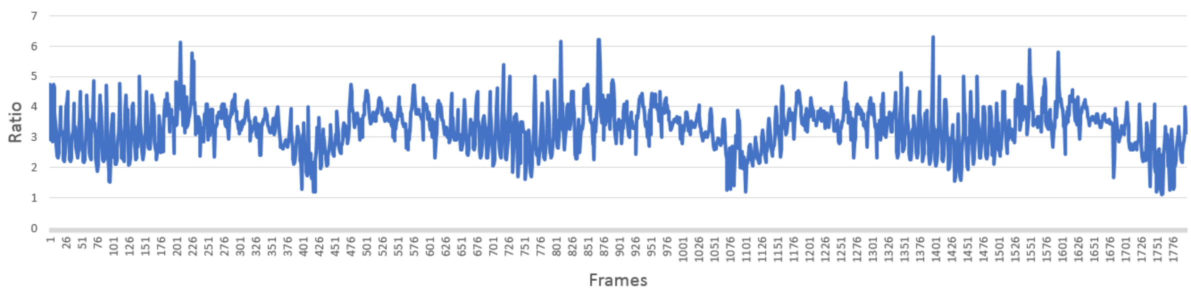
Figure 11: Extracted bounding box ratio for test 1, moving along a line of 20 metres length, turning at each end.



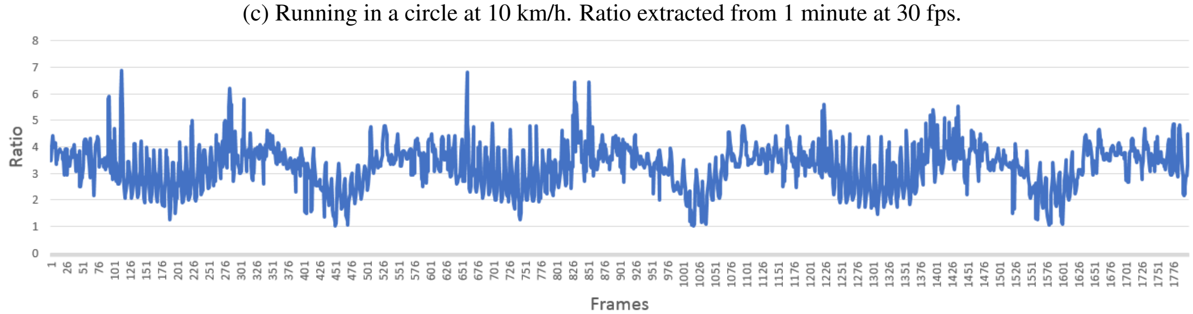
(a) Walking in a circle at 5 km/h. Ratio extracted from 1 minute at 30 fps.



(b) Running in a circle at 8 km/h. Ratio extracted from 1 minute at 30 fps.



(c) Running in a circle at 10 km/h. Ratio extracted from 1 minute at 30 fps.



(d) Running in a circle at 12 km/h. Ratio extracted from 1 minute at 30 fps.

Figure 12: Extracted bounding box ratio for test 2, moving along a circle of 20 metres diameter.

participants, future tests should also include higher running velocities to investigate how the correlation between step frequency and oxygen uptake extends, and test a possible need for additional features.

For this work we have focused on constant running velocities to be able to compare with measurements of oxygen uptake, which are usually measured after 2-3 minutes to ensure steady state conditions. Using this approach we aim to establish a correlation, after which we can start experimenting with dynamically changing velocities and different activity types. Eventually, we expect it to be possible to capture the dynamics of activities by analysing step frequency during few seconds of video. So, in addition to providing a non-invasive measurement method, visual analysis has the possibility of cutting down the response time for estimation of energy expenditure during sports.

7. Conclusion

We have here presented a new method for estimating energy expenditure of athletes from passive video data. The proposed method is based on automatic extraction of step frequency, independent of viewing angle. The presented results indicate a linear correlation between the new non-invasive measurement method and oxygen uptake, which is usually considered the ground truth. For future work a more extensive test should be performed to be able to statistically conclude on the correlation and identify areas for future research and improvements.

References

[1] Axis. Axis q1922, 2017. https://www.axis.com/hk/en/products/axis-q1922. 4

[2] S. Barris and C. Button. A review of vision-based motion analysis in sport. *Sports Medicine*, 38(12):1025–1043, 2008. 2

[3] CareFusion. Oxycon mobile, 2017. http://www.carefusion.com/documents/brochures/respiratory-care/cardiopulmonary/RC_Oxycon-Mobile-Device_BR_EN.pdf. 4

[4] A. Edgcomb and F. Vahid. Estimating daily energy expenditure from video for assistive monitoring. In *2013 IEEE International Conference on Healthcare Informatics*, pages 184–191, Sept 2013. 2

[5] B. W. Fudge, J. Wilson, C. Easton, L. Irwin, J. Clark, O. Haddow, B. Kayser, and Y. Pitsiladis. Estimation of oxygen uptake during fast running using accelerometry and heart rate. *Medicine and science in sports and exercise*, 39(1):192–198, 2007. 1

[6] R. Gade and T. B. Moeslund. Thermal cameras and applications: a survey. *Machine Vision and Applications*, 25(1):245–262, 2014. 2

[7] A. V. Hill and H. Lupton. The oxygen consumption during running. *The Journal of Physiology*, 56(supp):xxxii–xxxiii, 1922. 1

[8] N. Hongu, B. J. Orr, D. J. Roe, R. G. Reed, and S. B. Going. Global positioning system watches for estimating energy expenditure. *Journal of Strength and Conditioning Research*, 27(11):3216–3220, 2013. 1

[9] M. M. Jensen, M. K. Poulsen, T. Alldieck, R. G. Larsen, R. Gade, T. B. Moeslund, and J. Franch. Estimation of energy expenditure during treadmill exercise via thermal imaging. *Medicine and science in sports and exercise*, 48(12):2571–2579, 2016. 2

[10] M. Kim, J. Angermann, G. Bebis, and E. Folmer. Vizical: Accurate energy expenditure prediction for playing exergames. In *Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology*, UIST '13, pages 397–404, New York, NY, USA, 2013. ACM. 2

[11] A. Luke, K. C. Maki, N. Barkey, R. Cooper, and D. McGee. Simultaneous monitoring of heart rate and motion to assess energy expenditure. *Medicine and science in sports and exercise*, 29(1):144–148, 1997. 1

[12] 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. *Medicine and science in sports and exercise*, 42(1):170–178, 2010. 2

[13] R. Poppe. Vision-based human motion analysis: An overview. *Computer Vision and Image Understanding*, 108(1-2):4–18, Oct. 2007. 2

[14] L. Tao, T. Burghardt, M. Mirmehdi, D. Damen, A. Cooper, S. Hannuna, M. Camplani, A. Paiement, and I. Craddock. *Calorie Counter: RGB-Depth Visual Estimation of Energy Expenditure at Home*, pages 239–251. Springer International Publishing, 2017. 2

[15] G. Thomas, R. Gade, T. B. Moeslund, P. Carr, and A. Hilton. Computer vision for sports: current applications and research topics. *Computer Vision and Image Understanding*, In press, 2017. https://doi.org/10.1016/j.cviu.2017.04.011. 1

[16] P. F. Tsou and C. C. Wu. Estimation of calories consumption for aerobics using kinect based skeleton tracking. In *IEEE International Conference on Systems, Man, and Cybernetics*, pages 1221–1226, Oct 2015. 2

[17] L. Wang, W. Hu, and T. Tan. Recent developments in human motion analysis. *Pattern Recognition*, 36(3):585 – 601, 2003. 2

[18] Y. Yang, C. Liu, F. Tsow, D. Shao, H. Yu, S. Xia, and N. Tao. Remote quantification of workout energy expenditure with a cell phone camera. *IEEE Sensors Journal*, 16(23):8263–8270, Dec 2016. 2