

Facial video-based detection of physical fatigue for maximal muscle activity

ISSN 1751-9632

Received on 22nd July 2015

Revised 9th February 2016

Accepted on 8th March 2016

E-First on 22nd April 2016

doi: 10.1049/iet-cvi.2015.0215

www.ietdl.org

Mohammad A. Haque¹, Ramin Irani¹, Kamal Nasrollahi¹ ✉, Thomas B. Moeslund¹

¹Visual Analysis of People Laboratory, Aalborg University, Aalborg, Denmark

✉ E-mail: kn@create.aau.dk

Abstract: Physical fatigue reveals the health condition of a person at, for example, health checkup, fitness assessment, or rehabilitation training. This study presents an efficient non-contact system for detecting non-localised physical fatigue from maximal muscle activity using facial videos acquired in a realistic environment with natural lighting where subjects were allowed to voluntarily move their head, change their facial expression, and vary their pose. The proposed method utilises a facial feature point tracking method by combining a 'good feature to track' and a 'supervised descent method' to address the challenges that originate from realistic scenario. A face quality assessment system was also incorporated in the proposed system to reduce erroneous results by discarding low quality faces that occurred in a video sequence due to problems in realistic lighting, head motion, and pose variation. Experimental results show that the proposed system outperforms video-based existing system for physical fatigue detection.

1 Introduction

Fatigue is an important physiological parameter that usually describes the overall feeling of tiredness or weakness in human body. Fatigue may be either mental or physical or both [1]. Mental fatigue is a state of cortical deactivation due to prolonged periods of cognitive activity that reduces mental performance. On the other hand, physical fatigue refers to the declination of the ability of muscles to generate force. Stress, for example, makes people mentally exhausted, while hard work or extended physical exercise can exhaust people physically. Though, mental fatigue is related to cognitive activity, it can occur during a physical activity that comprises neurological phenomenon, for example directed attention as found in the area of intelligent transportation systems [2]. Unlike mental fatigue that is related to cognitive performance, physical fatigue specifically refers to muscles' inability to force optimally due to inadequate rest during a muscle activity [1]. Physical fatigue occurs from two types of activities: sub-maximal muscle activity (e.g. using a cycle ergometer or motor driven treadmill) and maximal muscle activity (e.g. pressing a dynamometer or lifting a load with great force) [3]. This kind of fatigue is a significant physiological parameter, especially for athletes or therapists. For example, by monitoring the occurrence of a patient's fatigue during physical exercise in rehabilitation scenarios, a therapist can change the exercise, make it easier or even stop it if necessary. Estimating the fatigue time offsets can also provide information in posterity health analysis [1].

A number of video-based non-invasive methods for fatigue detection and quantification have been proposed in the literature.

The methods utilised some features and clues, as shown in Fig. 1*a*, automatically extracted from a subject's facial video and discriminate between fatigue and non-fatigue classes automatically. For example, the works in [4] use eye blink rate and duration of eye closure for detection of fatigue occurrence due to sleep deprivation or directed attention. In addition to these features head pose and yawning behaviour in facial video is used in [5]. A review of facial video-based fatigue detection methods can be found in [2]. However, all of these methods address the detection of mental fatigue occurred from prolonged directed attention activity or sleep deprivation, more specifically known as driver fatigue, instead of physical fatigue occurred from maximal or sub-maximal muscle activity, as shown in Figs. 1*b* and *c*. Most of the available technologies for detecting physical fatigue occurrence (in terms of fatigue time offsets and/or fatigue level) use contact-based sensors such as force gauge, electromyogram (EMG), and mechanomyogram (MMG). Force gauge is minimally invasive, but it requires a device like a hand grip dynamometer [6]. EMG uses electrodes which require wearing adhesive gel patches [7]. MMG is based on an accelerometer or goniometer that requires direct skin contact and is sensitive to noise [8].

To the best of our knowledge, the only video-based non-invasive system for non-localised (i.e. not restricted to a particular muscle) physical fatigue detection in a maximal muscle activity scenario (as shown in Fig. 1*c*) is the one introduced in [9] which uses head-motion (shaking) behaviour due to fatigue in video captured by a simple webcam. It takes into account the fact that muscles start shaking when fatiguing contraction occurs in order to

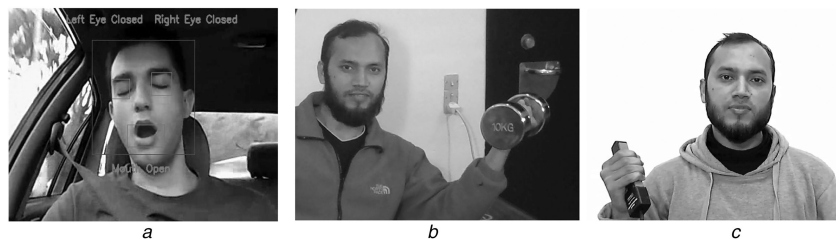


Fig. 1 Analysing facial video for different fatigue scenarios

a Driver's mental fatigue experiment (image taken from [5])

b Physical fatigue experiment using dumbbell

c Physical fatigue experiment using hand dynamometer

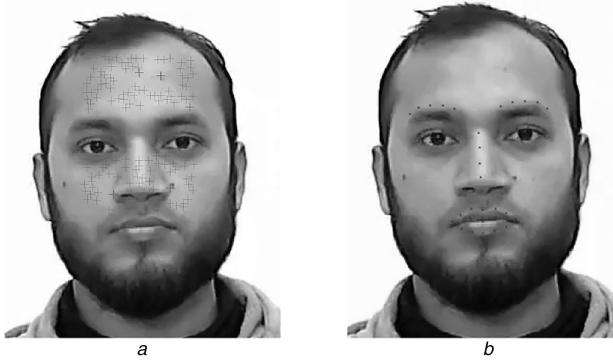


Fig. 2 Facial feature points and 49 facial landmarks in a face obtained by
a GFT-based tracking [9]
b SDM-based tracking [15]

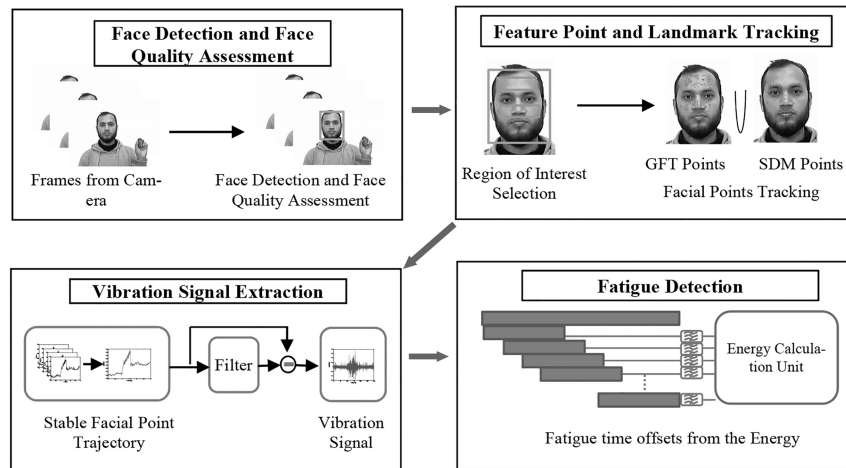


Fig. 3 Block diagram of the proposed system

send extra sensation signal to the brain to get enough force in a muscle activity and this shaking is reflected in the face. Inspired by [10] for heartbeat detection from ballistocardiogram, in [9] some feature points on the region of interest (ROI) (forehead and cheek) of the subject's face in a video are selected and tracked to generate trajectories of the facial feature points, and to calculate the energy of the vibration signal, which is used for measuring the onset and offset of fatigue occurrence in a non-localised notion. Though both physical fatigue and mental fatigue can occur simultaneously during a physical activity, the physiological mechanisms are not same. While mental fatigue represents the temporary reduction of cognitive performance, physical fatigue represents temporary reduction of force induced in muscle to accomplish a physical activity [2]. Unlike driver mental fatigue, physical fatigue for maximal muscle activity does not necessarily require a prolonged period. Thus, the visual clues found in the case of driver fatigue cannot be found in the case of physical fatigue from maximal muscle contraction. Changes in facial features in these two different types of fatigue are very different: in the driver mental fatigue eye blinking, yawning, varying head pose, and degree of mouth openness are used (as shown in Fig. 1*a*), while in the non-localised maximal muscle contraction-based physical fatigue *head-motion behaviour from shaking* is used. Consequently, physical fatigue occurred from maximal muscle activity cannot be detected or quantised by the computer vision methods used for detecting driver mental fatigue.

The previous facial video-based method in [9] for non-localised physical fatigue detection extracts some facial feature points as shown in Fig. 2*a*. Depending upon imaging scenario, the number of feature points and their position can vary. The method then employs signal processing techniques to detect head-motion trajectories from feature points in the video frames and estimates energy escalation to detect fatiguing contraction. However, the work in [9] assumes that there is neither internal facial motion, nor external movement or rotation of the head during the data

acquisition phase. We denote internal motion as facial expression and external motion as head pose. In real-life scenarios there are, of course, both internal and external head motions. The current method, therefore, fails due to an inability to detect and track the feature points in the presence of internal and external motion, and low texture in the facial region. Moreover, real-life scenarios challenge current methods due to low facial quality in video because of motion blur, bad posing, and poor lighting conditions [11]. The proposed system in this paper extends [9] by addressing the abovementioned shortcomings and thereby allows for automatic and more reliable detection of fatigue time offsets from facial video captured by a simple webcam. To address the shortcomings, we introduce a face quality assessment (FQA) method that prunes the captured video data so that low quality face frames cannot contribute to erroneous results [12, 13]. Following [10, 14], we track feature points (Fig. 2*a*) through a method good feature to track (GFT) with Kanady–Lucas–Tomasi (KLT) tracker, and then combine these trajectories with 49 facial landmark

trajectories (Fig. 2*b*), tracked by a supervised descent method (SDM) of [15, 16]. The idea of combining these two types of features has been developed in our paper [17], which was applied to heartbeat estimation from facial video. Here, we look at another application of this idea for physical fatigue estimation. The experiments are conducted on realistic datasets collected at the lab and a commercial fitness centre for fatigue measurement. The paper's contributions are as follows:

- We identify the limitations of the GFT-based tracking used in previous methods for physical fatigue detection and propose a solution using SDM-based tracking.
- We provide evidence for the necessity of combining the trajectories from the GFT and the SDM, instead of using the trajectories from either the GFT or the SDM.
- We introduce the notion of FQA in the physical fatigue detection context and demonstrate empirical evidence for its effectiveness.

The rest of the paper is organised as follows. Section 2 describes the proposed method. The results are summarised in Section 3. Finally, Section 4 concludes the paper.

2 Proposed method

The block diagram of the proposed method is shown in Fig. 3. The steps are explained in the following subsections.

2.1 Face detection and FQA

The first step of the proposed motion-based physical fatigue detection system is face detection from facial video, which has been accomplished by Viola and Jones [18] object detection framework using Haar-like features obtained from integral images.

Facial videos captured in real-life scenarios can be subject to the problems of pose variation, varying levels of brightness, and motion blur. When the intensity of these problems increases, the face quality decreases. A low quality face produces erroneous results in detecting facial features using either GFT or SDM [11]. To solve this problem, we pass the detected face to an FQA module. The FQA module assesses the quality of the face in the video frames. As investigated in [17, 19], four quality metrics can be critical for facial geometry analysis and detection of landmarks: resolution, brightness, sharpness, and out-of-plan face rotation (pose). Thus, the low quality faces can be discarded by calculating these quality metrics and employing thresholds to check whether the face needs to be discarded. The formulae to obtain these quality scores from a face are listed in [11]. Resolution score is calculated in terms of number of pixels, pose score is calculated by detecting the centre of mass in the binary image of the face region, sharpness is calculated by employing a low-pass filter to detect motion blur or unfocused capture, and brightness is calculated from the average of the illumination component of all the pixels in the face region. When we obtain the quality scores, following [17], we discard the low quality faces by the thresholds as follows: face resolution – 150×150 , brightness – 0.80, sharpness – 0.80, and pose – 0.20 by following [11]. As we detect the fatigue time offsets over a long video sequence (e.g. 30–180 s) for maximum muscle activity, discarding few frames (e.g. <5% of the total frames) does not affect much the regular characteristic of the trajectories, but removes the most erroneous segments coming from low quality faces. In fact, no frames are discarded if the quality score is not less than the thresholds. Missing points in the trajectory are removed by concatenating trajectory segments. The effect of employing FQA will be illustrated in the experimental evaluation section.

2.2 Feature points and landmarks tracking

As mentioned earlier, muscles start shaking when a subject becomes tired physically (the occurrence of physical fatigue from maximal muscle activity) [20]. The energy dispersed from this shaking is distinctively intense than the other types of head motion and is reflected in the face. Thus, physical fatigue can be determined from head motion by estimating the released shaking energy. Tracking facial feature points and generating trajectory help to record the head motion in facial video. This task was accomplished in [9] using merely a GFT-based method (utilises KLT tracker). In order to detect and track facial feature points in consecutive video frames, the GFT-based method uses an affine motion model to express changes in the intensity level in the face. It defines the similarity between two points in two frames using a so called ‘neighbourhood sense’ or window of pixels. Tracking a window of size $w_x \times w_y$ in the frame I to the frame J is defined on a point velocity parameter $\delta = [\delta_x \ \delta_y]^T$ for minimising a residual function f_{GFT} as follows:

$$f_{\text{GFT}}(\delta) = \sum_{x=p_x}^{p_x+w_x} \sum_{y=p_y}^{p_y+w_y} (I(x) - J(x+\delta))^2 \quad (1)$$

where $(I(x) - J(x+\delta))$ stands for $(I(x, y) - J(x+\delta_x, y+\delta_y))$, and $p = [p_x, p_y]^T$ is a point to track from the first frame to the second frame. According to the observations in [14], the quality of the estimate by this tracker depends on three factors: the size of the window, the texture of the image frame, and the amount of motion between frames. The GFT-based fatigue detection method assumes that the head does not have voluntary head motion during data capture. However, voluntary head motion (both external and internal) and low texture in facial videos are usual in real-life scenarios. Thus, the GFT-based tracking of facial feature points exhibits four problems. *First problem* arises due to low texture in the tracking window. This difficulty can be overcome by tracking feature points in corners or regions with high spatial frequency content, instead of forehead and cheek. *Second problem* arises by losing track in long video sequences due to point drifting in long video sequences. *Third problem* occurs in selecting an appropriate window size [i.e.

$w_x \times w_y$ in (1)]. If the window size is small, a deformation matrix to find the track is harder to estimate because the variations of motion within it are smaller and therefore less reliable. On the other hand, a bigger window is more likely to straddle a depth discontinuity in subsequent frames. *Fourth problem* comes when there is large optical flow in consecutive video frames. When there is voluntary motion or expression change in a face, the optical flow or face velocity in consecutive video frames is very high and GFT-based method misses the track due to occlusion [21]. Higher video frame rate may be able to address this problem, however this will require specialised camera instead of simple webcam. Due to these four problems, the GFT-based trajectory for fatigue detection leads to erroneous result in realistic scenarios where lighting changes and voluntary head motions exist.

A viable way to enable the GFT-based systems to detect physical fatigue in a realistic scenario is to track the facial landmarks by employing a face alignment system. Face alignment is considered as a mathematical optimisation problem and a number of methods have been proposed to solve this problem. The active appearance model (AAM) fitting [22] and its derivatives [23] were some of the early solutions in this area. The AAM fitting works by estimating parameters of an artificial model that is sufficiently close to the given image. In order to do that, AAM fitting was formulated as a Lukas–Kanade problem [24], which could be solved using Gauss–Newton optimisation [25]. A fast and effective solution to this was proposed recently in [15], which develops a SDM to minimise a non-linear least square function for face alignment. The SDM first uses a set of manually aligned faces as training samples to learn a mean face shape. This mean shape is then used as an initial point for an iterative minimisation of a non-linear least square function towards the best estimates of the positions of the landmarks in facial test images. The minimisation function can be defined as a function over Δx as

$$f_{\text{SDM}}(x_0 + \Delta x) = g(d(x_0 + \Delta x)) - \theta_{*,2}^2 \quad (2)$$

where x_0 is the initial configuration of the landmarks in a facial image, $d(x)$ indexes the landmarks configuration (x) in the image, g is a non-linear feature extractor, $\theta_* = g(d(x_*))$, and x_* is the configuration of the true landmarks. The scale invariant feature transform [11] is used as the feature extractor g . In the training images Δx and θ_* are known. By utilising these known parameters the SDM iteratively learns a sequence of generic descent directions, $\{\partial_n\}$, and a sequence of bias terms, $\{\beta_n\}$, to set the direction towards the true landmarks configuration x_* in the minimisation process, which are further applied in the alignment of unlabelled faces [15]. This working procedure of SDM in turn addresses the four previously mentioned problems of the GFT-based approach for head-motion trajectory extraction by as follows. *First*, the 49 facial landmark point tracked by SDM are taken only around eye, lip, and nose edges and corners, as shown in Fig. 2b. As these landmarks around the face patches have high spatial frequency and do not suffer from low texturedness, this eventually solves the problem of low texturedness. We cannot simply add these landmarks in the GFT-based tracking, because the GFT-based method has its own feature point selector. *Second*, SDM does not use any reference points in tracking. Instead, it detects each point around the edges and corners in the facial region of each video frame by using supervised descent directions and bias terms. Thus, the problems of point drifting do not occur in long videos. *Third*, SDM utilises the ‘neighbourhood sense’ on a pixel-by-pixel basis instead of a window. Therefore, window size is not relevant to SDM. *Fourth*, the use of supervised descent direction and bias terms allows the SDM to search selectively in a wider space and look after it from large optical flow problem. Thus, large optical flow cannot create occlusion in the SDM-based approach.

As in realistic scenarios the subjects are allowed to have voluntary head motion and facial expression change in addition to the natural cyclic motion, the GFT-based method results to either of the two consequences for videos having challenging scenarios: (i) completely missing the track of feature points and (ii) erroneous tracking. We observed more than 80% loss of feature points by the

Table 1 Behaviour of the GFT, SDM, and a combination of both methods for facial points tracking in different scenarios

Scenario	Challenge	GFT	SDM	Combination
low texture in video	number of facial points available to effectively generating motion trajectory	<i>bad</i>	<i>good</i>	<i>better</i>
long video sequence	facial point drifting during tracking	<i>bad</i>	<i>good</i>	<i>better</i>
appearance of voluntary head motion in video	optical occlusion and depth discontinuity of window based tracking	<i>bad</i>	<i>good</i>	<i>better</i>
perfect scenario	none of the aforementioned challenges	<i>good</i>	<i>good</i>	<i>better</i>

system in such cases. The GFT-based method, in fact, fails to preserve enough information to estimate fatigue from trajectories even though the video have minor expression change or head motion voluntarily. On the other hand, the SDM does not miss or erroneously track the landmarks in the presence of voluntary facial motions or expression change or low texturedness as long as the face is qualified by the FQA. Thus, the system can find enough trajectories to detect fatigue. However, the GFT-based method uses a large number of facial points to track when compared with SDM. This matter causes the GFT-based method to generate a better trajectory than SDM when there is no voluntary motion or low texturedness. Following the above discussions Table 1 summarises the behaviour of GFT, SDM, and a combination of these two methods in facial point tracking. We observe that a combination of trajectories obtained by GFT and SDM-based methods can produce better results in cases where subjects may have both motion and non-motion periods. We thus propose to combine the trajectories. In order to generate combined trajectories, the face is passed to the GFT-based tracker to generate trajectories from facial feature points and then appended with the SDM trajectories.

2.3 Vibration signal extraction

To obtain vibration signal for fatigue detection, we take the average of all the trajectories obtained from both feature and landmark points of a video by as follows:

$$T(n) = \frac{1}{M} \sum_{m=1}^M (y_m(n) - \bar{y}_m) \quad (3)$$

where $T(n)$ is the shifted mean filtered trajectory, $y_m(n)$ is the n th frame of the trajectory m , M is the number of the trajectories, N is the number of the frames in each trajectory, and \bar{y}_m is the mean value of the trajectory m given by

$$\bar{y}_m = \frac{1}{N} \sum_{n=1}^N y_m(n) \quad (4)$$

The vibration signal that keeps the shaking information is calculated from T by using a window of size R by

$$V_s(n) = T(n) - \frac{1}{R} \sum_{r=0}^{R-1} T(n-r) \quad (5)$$

The obtained signal is then passed to the fatigue detection block.

2.4 Physical fatigue detection

To detect the released energy of the muscles reflected in head shaking we need to segment the vibration signal V_s from (5) with an interval of Δt_{sec} . Segmenting the signal V_s helps detecting the fatigue in temporal dimension. After windowing, each block is filtered by a passband ideal filter. Fig. 4a shows the power of the

filtered vibrating signal with a cut-off frequency interval of [3–5] hertz (Hz). The cut-off frequency was determined empirically in [9]. We observe that the power of the signal rises when fatigue happens in the interval of 16.3–40.6 s in this figure. After filtering, the energy of i th block is calculated as

$$E_i = \sum_{j=1}^M |Y_{ij}|^2 \quad (6)$$

where E_i is the calculated energy of the i th block, Y_{ij} is the fast Fourier transform (FFT) of the signal $V_{s,i}$, and M is the length of Y . Finally, fatigue occurrence is detected by

$$F_i = k \frac{E_i}{(1/N) \sum_{j=1}^N E_j} \tanh \left(\gamma \left(\frac{E_i}{(1/N) \sum_{j=1}^N E_j} - 1 \right) \right) \quad (7)$$

where F_i is the fatigue index, N is the number of the initial blocks in the normal case (before starting the fatigue), K is the amplitude factor, and γ is a slope factor. Experimentally, we obtained reasonable results with $k=10$ and $\gamma=0.01$. As observed in [9], employing a bipolar sigmoid (tangent hyperbolic) function to E_i in (7) suppresses the noise peaks out of fatigue region that appear in the results because of the facial expression and/or the voluntary motion. Figs. 4b and c illustrate the effect of the sigmoid function on the output results and Figs. 4d and e depict the effect in values. To realise the effect of such noise suppression in percentage, by following [26] we use the following metric:

$$\text{SUP}_i = \frac{F_i}{F_{\max}} \times 100\% \quad (8)$$

where SUP_i is the ratio of the noise to the released fatigue energy. If we employ (8) on Figs. 4d and e, we obtain values 8.94%, 11.65% and 0.77%, 1.38%, respectively, for the noise datatips shown in the figures. It can be noticed that before employing the suppression the noise to fatigue energy were $\sim 10\%$, however reduced to $\sim 1\%$ after employing the suppression. When we obtain the fatigue index, the starting and ending time of fatigue occurrence in a subject's video are detected by employing a threshold with value: 1.0 to the normalised fatigue index, as the bipolar sigmoid suppresses the signal energy out of fatigue region to <1.0 by (7). Fatigue starts when the fatigue index exceeds the threshold upward and fatigue ends when fatigue index exceeds the threshold downward.

3 Experimental results

3.1 Experimental environment

The proposed method was implemented using a combination of Matlab (2013a) and C++ environments. We integrated the SDM [15] with the GFT-based tracker from [9, 11] to develop the system as explained in the methodology section. We collected four experimental video databases to generate results: a database for demonstrating the effect of FQA, a database with voluntary motions in some moments for evaluating the performance of GFT, SDM, and the combination of GFT and SDM, a database collected from the subjects in a natural laboratory environment, and a database collected from the subjects at a real-life environment in a commercial fitness centre. We named the databases as 'FQA_Fatigue_Data', 'Eval_Fatigue_Data', 'Lab_Fatigue_Data', and 'FC_Fatigue_Data'. All the video clips were captured in VGA resolution using a Logitech C310 webcam. The videos were collected from 16 subjects (including both male and female from different ethnicities with the ages between 25 and 40 years) after adequately informing the subjects about the concepts of maximal muscle fatigue and the experimental scenarios. Subjects exposed their face in front of the cameras while performing maximal muscle activity by using a handgrip dynamometer for about 30–180 s (varies from subject to subject). Subjects were free to have natural head motion and expression variation due to activity prompted by using the dynamometer. Both setups (in the laboratory

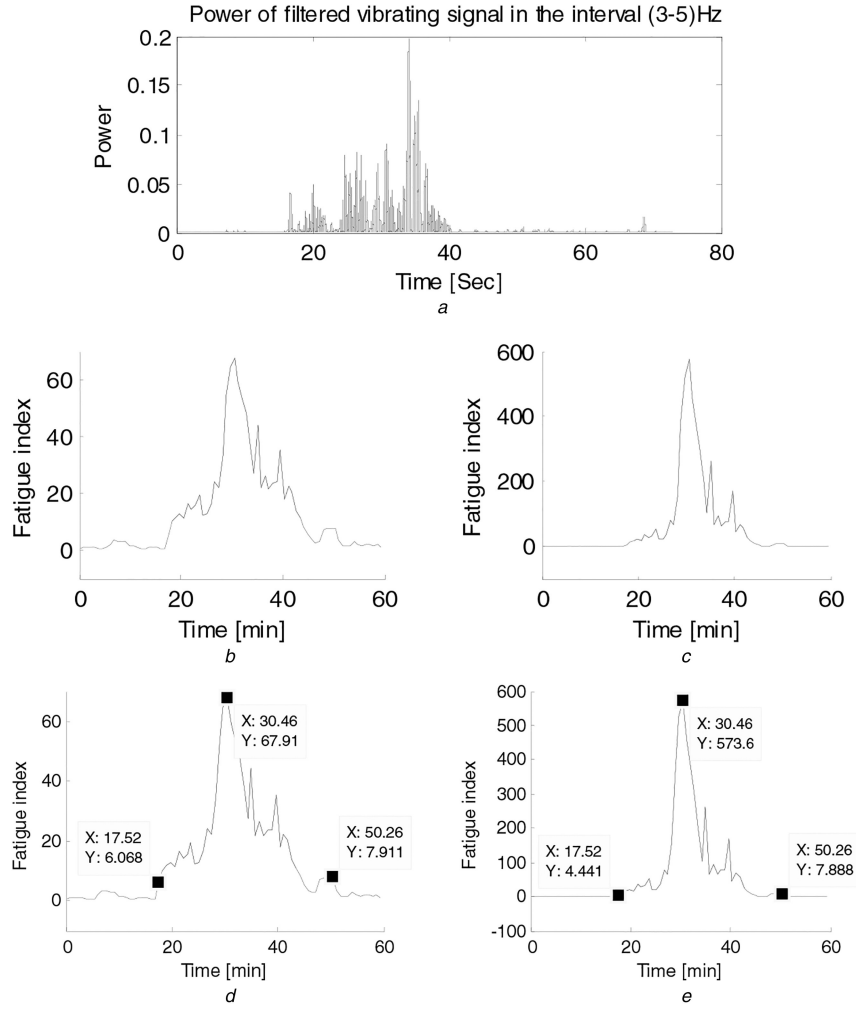


Fig. 4 Analysing trajectory for fatigue detection

a Power of the trajectory where the blue region is the resting time and the red region shows the fatigue due to exercise in the interval 16.3–40.6 s
b and *c* Before and after using a bipolar sigmoid function to suppress the noise peaks, respectively
d and *e* Depicts the effect of bipolar sigmoid in values corresponding to (*b*) and (*c*), respectively

and in the fitness centre) used indoor lighting for video capturing and the dynamometer reading to measure ground truth for fatigue. The FQA_Fatigue_Data has 12 videos, each of which contains some low quality face in some moments. The Eval_Fatigue_Data has 17 videos with voluntary motion, Lab_Fatigue_Data has 54 videos and the FC_Fatigue_Data has 11 videos in natural scenario.

As physical fatigue in a video clip occurs between a starting time and an ending time, the starting and ending times detected from the video by the experimental methods should match with the starting and ending times of fatigue obtained from the ground truth dynamometer data. Thus, we analysed and measured the error between the ground truth and the output of the experimental methods for starting and ending time agreement by defining a parameter μ . This parameter expresses the average of the total of starting and ending point distances of fatigue occurrence for each subject in the datasets, and is calculated as follows:

$$\mu = \frac{1}{n} \sum_{i=1}^n (|G_s^i - R_s^i| + |G_e^i - R_e^i|) \quad (9)$$

where n is the number of video (subjects) in a dataset, G_s^i is the ground truth of the starting point of fatigue, G_e^i is the ground truth of the ending point of fatigue, R_s^i is the calculated starting point of fatigue, and R_e^i is the calculated ending point of fatigue.

3.2 Performance evaluation

The proposed method used a combination of the SDM- and GFT-based approaches for trajectory generation from the facial points.

Fig. 5 shows the calculated average trajectories of tracked points in two experimental videos. We depicted the trajectories obtained from GFT-based tracker, SDM and another recent face alignment algorithm Par-CLR [27] for two facial videos with voluntary head motion. As observed from the figure, the GFT and SDM-based trackers provide similar trajectories when there is little head motion (video1, first row of Fig. 5). On the other hand, Par-CLR provides a trajectory very different than the other two because of tracking on false positive face in the video frames. When the voluntary head motion is sizable (beginning of video 2, second row of Fig. 6), GFT-based method fails to track the point accurately and thus produces an erroneous trajectory. However, SDM provides stable trajectory in this case. Thus, lack in proper selection of method(s) for trajectory generation can contribute to erroneous results in estimating fatigue time offsets, as we observe for the GFT-based tracker and the recently proposed Par-CLR in comparison to SDM.

For the fatigue time offset measurement experiment we asked the test subjects to squeeze the handgrip dynamometer as much as they could. As they did this we recorded their face. The squeezed dynamometer provides a pressure force, which is used as the ground truth data in fatigue detection. Fig. 6a displays the data recorded while using the dynamometer, where the part of the graph with a falling force indicates the fatigue region. The measured fatigue level from the dynamometer reading is shown in Fig. 6b. Fatigue in this figure happens when the fatigue level sharply goes beyond a threshold defined in [9]. Comparative experimental results for fatigue detection using different methods are shown in the next section.

We conducted experiments to evaluate the effect of employing FQA, and a combination of GFT and SDM in the proposed system.

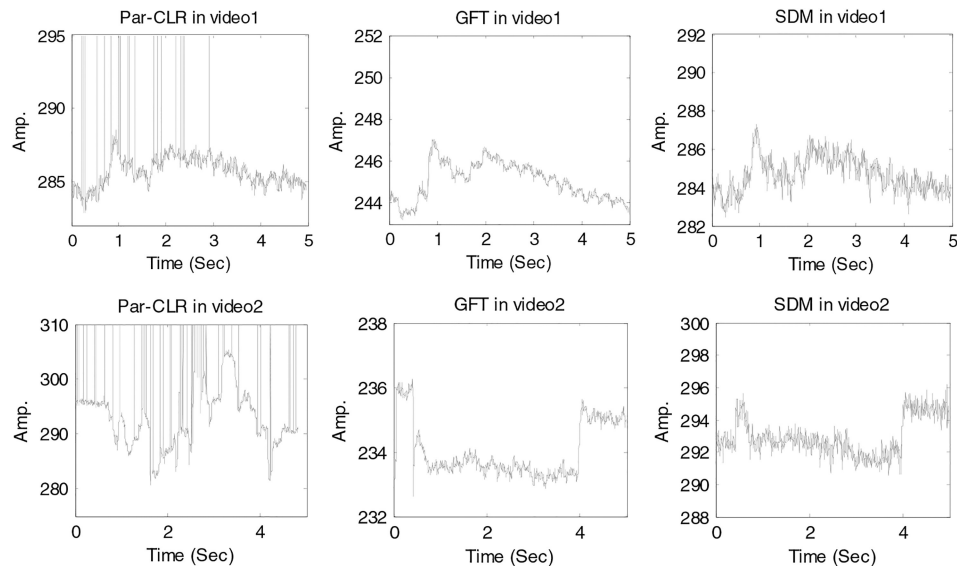


Fig. 5 Trajectories of tracking points extracted by Par-CLR [27], GFT [14], and SDM [15] from 5 s of two experimental video sequences with continuous small motion (for video1 in the first row) and large motion at the beginning and end (for video2 in the second row)

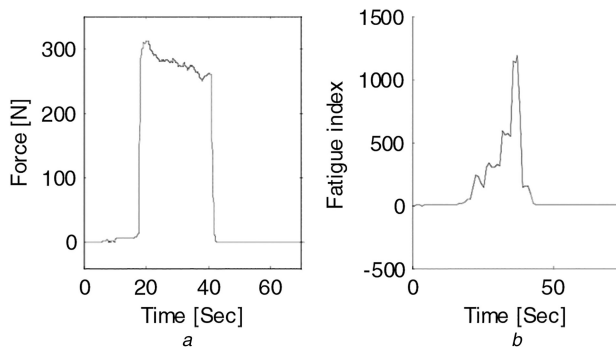


Fig. 6 Detection of physical fatigue due to maximal muscle activity
a Dynamometer reading during fatigue event
b Fatigue time spectral map for fatigue time offset measurement. The blue region is the resting time and the red region shows the fatigue due to exercise

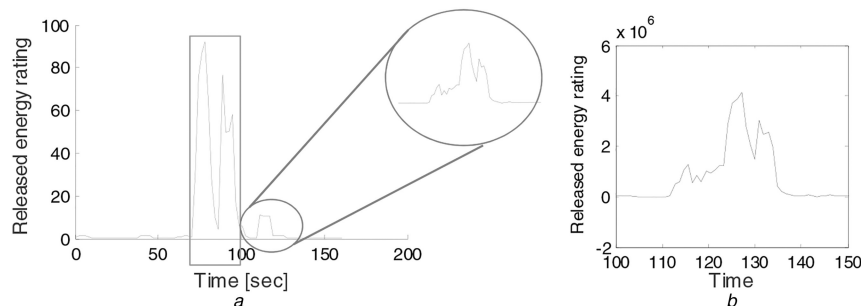


Fig. 7 Analysing the effect of employing FQA on a trajectory obtained from an experimental video
a Without employing FQA (red circle presents the real fatigue location and green rectangle presents the moments of low quality faces)
b Employing FQA [presenting the area within the red circle of (a)]

Fig. 7 shows the effect of employing FQA on a trajectory obtained from a subject's video. It is observed that low quality face region (due to pose variation) shows erroneous trajectory and contributed to the wrong detection of fatigue onset (Fig. 7*a*). When, FQA module discarded this region, the actual fatigue region was detected as shown in Fig. 7*b*. Table 2 shows the results of employing FQA on the FQA_Fatigue_Data, and evaluating the performance of GFT, SDM, and the combination of these two on the Eval_Fatigue_Data. From the results it is observed that when videos have low quality faces (which are true for all the videos of the FQA_Fatigue_Data), automatic detection of fatigue time stamps exhibited very high error due to wrong place of detection. When we employed FQA the fatigue was detected in the expected time with minor error. While comparing GFT, SDM, and the

combination of these two, we observe that the SDM minimally outperformed the GFT, however the combination worked better. These observations came with the agreement of the characteristics we listed in Table 1.

3.3 Performance comparison

To the best of our knowledge, the method of [9] is the first and the only work in the literature to detect physical fatigue from facial video. Other methods for facial video-based fatigue detection work for driver mental fatigue [2], and use different scenarios than what is used in physical fatigue detection environment. Thus, we have compared the performance of the proposed method merely against the method of [9] on the experimental datasets. Fig. 8*a* shows the physical fatigue detection duration for a subset of database Lab_Fatigue_Data in a bar diagram. The height of the bar shows the duration of fatigue in seconds. Fig. 8*b* shows the total detection error in seconds for starting and ending points of fatigue in the

videos. From the result it is observed that the proposed method detected the presence of fatigue (expressed by fatigue duration) more accurately than the previous method of [9] in comparison to the ground truth as shown in Fig. 8*a* and demonstrated better agreement with the starting and ending time of fatigue with the ground truth as shown Fig. 8*b*. Table 3 shows the fatigue detection results on both Lab_Fatigue_Data and FC_Fatigue_Data, and compares the performance between the state of the art method of [9] and the proposed method. While analysing the agreement with the starting and ending time of fatigue with the ground truth, we observed that the proposed method shows more consistency than the method of [9] both in the Lab_Fatigue_Data experimental scenario and the FC_Fatigue_Data real-life scenario. However, the performance is higher for Lab_Fatigue_Data than

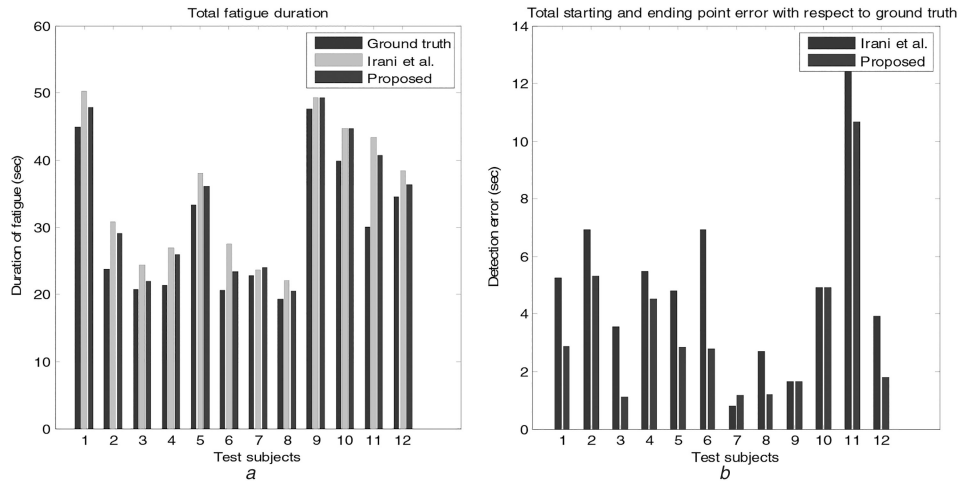


Fig. 8 Comparison of physical fatigue detection results of the proposed method with the Irani's method [9] on a subset of the Lab_Fatigue_Data

a Total duration of fatigue

b Total starting and ending point error in detection

Table 2 Analysing the effect of the FQA and evaluating the performance of GFT, SDM, and the combination of GFT and SDM in fatigue detection

Dataset	Scenario	Average of the total of the starting and ending point distance of fatigue occurrence for each subject in a dataset, μ
FQA_Fatigue_Data	without FQA	65.32
	with FQA	3.79
Eval_Fatigue Data	GFT	6.81
	SDM	6.35
	combination of GFT and SDM	5.16

Table 3 Performance comparison between the proposed method and a state-of-the-art method of physical fatigue detection on experimental datasets

No	Dataset name	Average of the total of the starting and ending point distance of fatigue occurrence for each subject in a dataset, μ	Irani et al. [9]	The proposed method
1.	Lab_Fatigue_Data		7.11	4.59
2.	FC_Fatigue_Data		3.35	2.65

FC_Fatigue_Data. We believe the realistic scenario of a commercial fitness centre (in terms of lighting and subject's natural behaviour) contributes to lower performance. The computational time of the proposed method suggest that the method is doable for real-time application, because it requires only 3.5 ms (app.) processing time for each video frame in a platform with 3.3 GHz processor and 8 GB RAM.

4 Conclusions

This paper proposes a physical fatigue detecting system from facial video captured by a simple webcam. The proposed system overcomes the drawbacks of previous facial video-based method of [9] by extending the application of SDM over GFT-based tracking and employing FQA. The previous method works well only when there is neither voluntary motion of the face nor change of expression, and when the lighting conditions help keeping sufficient texture in the forehead and cheek. The proposed method overcomes these problems by using an alternative facial landmarks tracking system (the SDM-based system) along with the previous feature points tracking system (the GFT-based system) and

provides competent results. The performance of the proposed system showed very high accuracy in proximity to the ground truth not only in a laboratory setting with controlled environment, as considered in [9], but also in a real-life environment in a fitness centre where faces have some voluntary motion or expression change and lighting conditions are normal.

The proposed method has some limitations. The camera was placed in close proximity to the face (about 1 m away) because the GFT-based feature tracker in the combined system does not work well if the face is far from the camera during video capture. Moreover, the fatigue detection of the proposed system does not take into account the sub-maximal muscle activity due to lack of reliable ground truth data for fatigue from sub-maximal muscle activity. Future work should address these points.

5 References

- [1] Watanabe, Y., Evengard, B., Natelson, B.H., et al.: 'Fatigue science for human health' (Springer Science & Business Media, 2007)
- [2] Sigari, M.H., Pourshahabi, M.R., Soryani, M., et al.: 'A review on driver face monitoring systems for fatigue and distraction detection', *Int. J. Adv. Sci. Technol.*, 2014, **64**, pp. 73–100
- [3] Baptista, R.R., Scheeren, E.M., Macintosh, B.R., et al.: 'Low-frequency fatigue at maximal and submaximal muscle contractions', *Braz. J. Med. Biol. Res.*, 2009, **42**, (4), pp. 380–385
- [4] Alioua, N., Amine, A., Rziza, M.: 'Driver's fatigue detection based on yawning extraction', *Int. J. Veh. Technol.*, 2014, **2014**, pp. 1–7
- [5] Sacco, M., Farrugia, R.A.: 'Driver fatigue monitoring system using support vector machines'. 2012 Fifth Int. Symp. on Communications Control and Signal Processing (ISCCSP), 2012, pp. 1–5
- [6] McArdle, W.D., Katch, F.I.: 'Essential exercise physiology' (Lippincott Williams and Wilkins, Philadelphia, 2010, 4th edn.)
- [7] Stoykov, N.S., Lowery, M.M., Kuiken, T.A.: 'A finite-element analysis of the effect of muscle insulation and shielding on the surface EMG signal', *IEEE Trans. Biomed. Eng.*, 2005, **52**, (1), pp. 117–121
- [8] Raez, M.B.I., Hussain, M.S., Mohd-Yasin, F.: 'Techniques of EMG signal analysis: detection, processing, classification and applications', *Biol. Proced. Online*, 2006, **8**, pp. 11–35
- [9] Irani, R., Nasrollahi, K., Moeslund, T.B.: 'Contactless measurement of muscles fatigue by tracking facial feature points in a video'. IEEE Int. Conf. on Image Processing (ICIP), 2014, pp. 1–5
- [10] Balakrishnan, G., Durand, F., Guttag, J.: 'Detecting pulse from head motions in video'. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), 2013, pp. 3430–3437
- [11] Haque, M.A., Nasrollahi, K., Moeslund, T.B.: 'Quality-aware estimation of facial landmarks in video sequences'. IEEE Winter Conf. on Applications of Computer Vision (WACV), 2015, pp. 1–8
- [12] Haque, M.A., Nasrollahi, K., Moeslund, T.B.: 'Real-time acquisition of high quality face sequences from an active pan-tilt-zoom camera'. Tenth IEEE Int. Conf. on Advanced Video and Signal Based Surveillance (AVSS), 2013, pp. 443–448
- [13] Klonovs, J., Haque, M.A., Krueger, V., et al.: 'Distributed computing and monitoring technologies for older patients' (Springer International Publishing, 2015, 1st edn.)
- [14] Shi, J., Tomasi, C.: 'Good features to track'. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), 1994, pp. 593–600
- [15] Xiong, X., De la Torre, F.: 'Supervised descent method and its applications to face alignment'. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), 2013, pp. 532–539

- [16] Tzimiropoulos, G., Pantic, M.: 'Optimization problems for fast AAM fitting in-the-wild'. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), 2013, pp. 593–600
- [17] Haque, M.A., Irani, R., Nasrollahi, K., *et al.*: 'Heartbeat rate measurement from facial video (accepted)', *IEEE Intell. Syst.*, 2015
- [18] Viola, P., Jones, M.J.: 'Robust real-time face detection', *Int. J. Comput. Vis.*, 2004, **57**, (2), pp. 137–154
- [19] Haque, M.A., Nasrollahi, K., Moeslund, T.B.: 'Constructing facial expression log from video sequences using face quality assessment'. Ninth Int. Conf. on Computer Vision Theory and Applications (VISAPP), 2014, pp. 1–8
- [20] Young, R.R., Hagbarth, K.-E.: 'Physiological tremor enhanced by manoeuvres affecting the segmental stretch reflex', *J. Neurol. Neurosurg. Psychiatry*, 1980, **43**, pp. 248–256
- [21] Bouguet, J.: 'Pyramidal implementation of the Lucas Kanade feature tracker', *Intel Corp. Microprocess. Res. Labs*, 2000
- [22] Cootes, T.F., Edwards, G.J., Taylor, C.J.: 'Active appearance models', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2001, **23**, (6), pp. 681–685
- [23] Batur, A.U., Hayes, M.H.: 'Adaptive active appearance models', *IEEE Trans. Image Process.*, 2005, **14**, (11), pp. 1707–1721
- [24] Baker, S., Matthews, I.: 'Lucas–Kanade 20 years on: a unifying framework', *Int. J. Comput. Vis.*, 2004, **56**, (3), pp. 221–255
- [25] Lucey, S., Navarathna, R., Ashraf, A.B., *et al.*: 'Fourier Lucas–Kanade algorithm', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2013, **35**, (6), pp. 1383–1396
- [26] Nasrollahi, K., Moeslund, T.B., Rahmati, M.: 'Summarization of surveillance video sequences using face quality assessment', *Int. J. Image Graph.*, 2011, **11**, (02), pp. 207–233
- [27] Asthana, A., Zafeiriou, S., Cheng, S., *et al.*: 'Incremental face alignment in the wild'. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), 2014, pp. 1–8