Aalborg Universitet

AALBORG UNIVERSITY

DENMARK

Heartbeat Rate Measurement from Facial Video

Haque, Mohammad Ahsanul; Irani, Ramin; Nasrollahi, Kamal; Moeslund, Thomas B.

Heartbeat Rate Measurement from Facial Video

Mohammad A. Haque, Ramin Irani, Kamal Nasrollahi, and Thomas B. Moeslund,
Aalborg University

Heartbeat rate (HR) is an important physiological parameter that provides information about the condition of the human body's cardiovascular system in applications such as medical diagnosis, rehabilitation training programs, and fitness assessments.[1] Increasing or decreasing a patient's HR

An effective approach for measuring heartbeat rate from facial video uses a facial feature point-tracking method that combines a good feature to track method with a supervised descent method to overcome limitations in current systems.

beyond the norm in a fitness assessment or during rehabilitation training, for example, can show how exercise affects the trainee and can indicate whether continuing the exercise is safe.

HR is typically measured by an electrocardiogram (ECG) reading via sensors placed on the body. A recent study noted that blood circulation causes periodic subtle changes to facial skin color,[2] a fact utilized for HR estimation[3–7] and applications of heartbeat signal from facial video in other works.[8–10] These facial color-based methods, however, aren't effective when taking into account the sensitivity to color noise and changes in illumination during tracking. Thus, G. Balakrishnan and col-leagues proposed a system for measuring HR based on the fact that the flow of blood through the aorta causes invisible motion in the head (which can be observed via ballistocardiography) due to pulsation of the heart muscles[11] (and an improvement of this method appears elsewhere[12]). These motion-based methods[11,12] extract facial feature points from the forehead and cheek (as shown in Figure 1a)

by a method called good feature to track (GFT). They then employ the Kanade-Lucas-Tomasi (KLT) feature tracker[13] to generate the motion trajectories of feature points and some signal-processing methods to estimate cyclic head motion frequency as the subject's HR. These calculations are based on the assumption that the head is (or is nearly) static during facial video capture, meaning there's neither internal facial motion nor external movement 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. Current methods, therefore, fail due to their inability to detect and track feature points in the presence of internal and external motion as well as low texture in the facial region. Moreover, real-life scenarios challenge current methods due to low facial quality in video caused by motion blur, bad posing, and poor lighting conditions.[14] These low-quality facial frames induce noise in the motion trajectories obtained for measuring HR.

Our proposed system addresses the aforementioned shortcomings and advances current automatic systems for reliably measuring HR. We introduce a face quality assessment (FQA) method that prunes the captured video data so that low-quality face frames can't contribute to erroneous results.[15,16] We then extract GFT feature points and combine them with facial landmarks (Figure 1b) extracted via the supervised descent method (SDM).[17] A combination of these two methods for vibration signal generation lets us obtain stable trajectories that, in turn, allow for a better estimation of HR. We conducted our experiments on a publicly available database and on a local database collected at the lab and at a commercial fitness center. Our experimental results show that our system outperforms state-of-the-art systems for HR measurement.

Theory

Tracking facial feature points to detect head motion in consecutive facial video frames has been accomplished[11,12] by using the GFT-based method, which uses an affine motion model to express changes in the face's level of intensity. Tracking a window of size $w_x \times w_y$ in frame I to frame J is defined on a point velocity parameter $\delta = (\delta_x \delta_y)^T$ for minimizing a residual function f_{GFT} that's defined by

$$f_{GFT}(\delta) = \sum_{x=p_x}^{p_x+w_x} \sum_{y=p_y}^{p_y+w_y} \left(I(\mathbf{x}) - J(\mathbf{x}+\delta) \right)^2,$$

(1)

where $(I(\mathbf{x}) - J(\mathbf{x} + \delta))$ stands for $(I(x, y) - J(x + \delta_x, y + \delta_y))$, and $\mathbf{p} = (p_x, p_y)^T$ is a point to track from the first frame to the second. According to observations,[18] the quality of the estimate by this tracker depends on three factors: window size, image
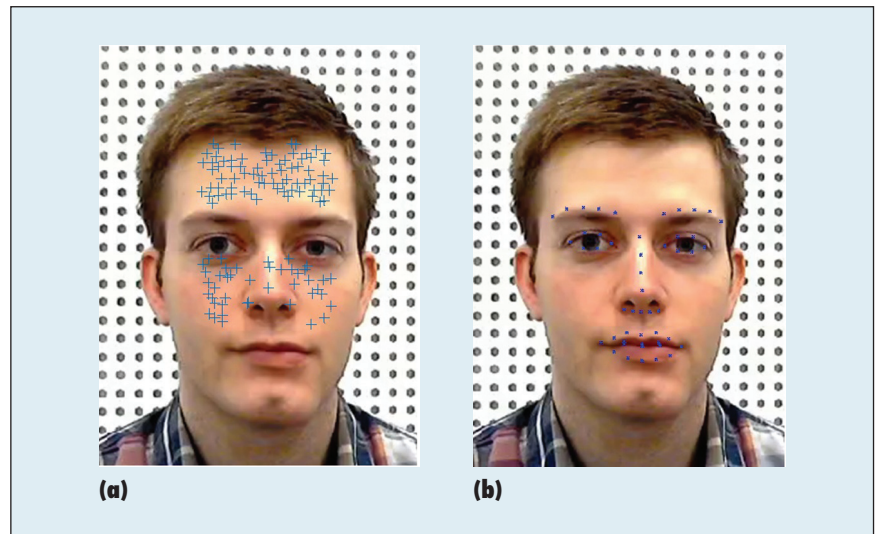

Figure 1. Different facial feature tracking methods: (a) facial feature points extracted by the good feature to track method and (b) facial landmarks obtained by the supervised descent method. While GFT extracts a large number of points, SDM merely uses 49 predefined points to track.

frame texture, and motion between frames. Thus, in the presence of voluntary head motion (both external and internal) and low texture in facial videos, the GFT-based tracking exhibits the following problems:

- *Low texture in the tracking window.* In general, not all parts of a video frame contain complete motion information because of an aperture problem. This difficulty can be overcome by tracking feature points in corners or regions with high spatial frequency content. However, GFT-based systems for HR use the feature points from the forehead and cheek that have low spatial frequency content.
- *Lost tracking in a long video sequence.* The GFT-based method applies a threshold to the cost function $f_{GFT}(\delta)$ to declare a point "lost" if the cost function is higher than the threshold. While tracking a point over many frames of a video,[11,12] the point can drift throughout the extended sequences and could be prematurely declared "lost."
- *Window size.* When the window size (such as $w_x \times w_y$ in Equation 1)

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.

- *Large optical flow vectors in consecutive video frames.* When there's voluntary motion or expression change in a face, the optical flow or face velocity in consecutive video frames is very high, and the GFT-based method misses the track due to occlusion.[13]

Instead of tracking feature points with the GFT-based method, facial landmarks can be tracked by employing a face alignment system. The Active Appearance Model (AAM) fitting[19] and its derivatives[20] are some of the early solutions for face alignment. A fast and highly accurate AAM fitting approach that was proposed recently[17] is SDM, which uses a set of manually aligned faces as training samples to learn a mean face shape. This shape is then used as an initial point for an iterative minimization of a nonlinear least square function toward the best estimates of the positions of the landmarks in facial

test images. The minimization function can be defined as a function over Δx:

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

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 nonlinear feature extractor, $\theta_* = g(d(x_*))$, and x_* is the configuration of the true landmarks. In the training images, Δx and θ_* are known. By utilizing these known parameters, SDM iteratively learns a sequence of generic descent directions, $\{\partial_n\}$, and a sequence of bias terms, $\{\beta_n\}$, to set the direction of search toward the true landmarks configuration x_* in the minimization process, which are further applied in the alignment of unlabeled faces.[17] The evaluation of the descent directions and bias terms is accomplished by

$$x_n = x_{n-1} + \partial_{n-1}\sigma(x_{n-1}) + \beta_{n-1}, \tag{3}$$

where $\sigma(x_{n-1}) = g(d(x_{n-1}))$ is the feature vector extracted at the previous landmark location x_{n-1}, x_n is the new location, and ∂_{n-1} and β_{n-1} are defined as

$$\partial_{n-1} = -2 \times \mathbf{H}^{-1}(x_{n-1}) \times \mathbf{J}^{T}(x_{n-1}) \times g\big(d(x_{n-1})\big) \tag{4}$$

and

$$\beta_{n-1} = -2 \times \mathbf{H}^{-1}(x_{n-1}) \times \mathbf{J}^{T}(x_{n-1}) \times g\big(d(x_*)\big), \tag{5}$$

where $\mathbf{H}(x_{n-1})$ and $\mathbf{J}(x_{n-1})$ are, respectively, the Hessian and Jacobian matrices of the function g evaluated at (x_{n-1}). The succession of x_n converges to x_* for all images in the training set.

SDM is free from the problems of the GFT-based tracking approach for the following reasons:

- *Low texture in the tracking window.* SDM's 49 facial landmarks are taken from face patches around the eye, lip, and nose (as shown in Figure 1b), which have high spatial frequency due to the existence of edges and corners.[18]
- *Lost tracking in a long video sequence.* SDM doesn't 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 as shown in Equations 3, 4, and 5. Thus, the problems of point drifting or dropping a point too early don't occur.
- *Window size.* SDM doesn't define the facial landmarks by using the window-based "neighborhood sense" and, thus, doesn't use a window-based point tracking system. Instead, SDM utilizes the neighborhood sense on a pixel-by-pixel basis along with the descent detections and bias terms.
- *Large optical flow vectors in consecutive video frames.* As mentioned elsewhere,[13] occlusion can occur by large optical flow vectors in consecutive video frames. As a video with human motion satisfies temporal stability constraint,[21] increasing the search space can be a solution. SDM uses supervised descent direction and bias terms that allow for selective searching in a wider space with high computational efficiency.

Although the GFT-based method fails to preserve enough information to measure HR when the video indicates facial expression change or head motion, it uses a larger number of facial feature points (more than 150) to track than SDM (only 49 points). This causes the GFT-based method to generate a better trajectory than SDM when there's no voluntary motion. On the other hand, SDM doesn't miss or erroneously track landmarks in the presence of voluntary facial motions. To exploit the advantages of both methods, we can use a combination of GFT- and SDM-based tracking outcomes. Merely using GFT or SDM to extract facial points in cases where subjects could have both voluntary motion and nonmotion periods doesn't produce competent results.

The Proposed Method

We follow a four-step procedure from facial video acquisition to HR calculation in this article. Figure 2 shows a block diagram of our proposed method, which is described in the following subsections.

Face Detection and Face Quality Assessment

The first step of the proposed motion-based system is face detection from facial video acquired by a webcam. We employed the Haar-like features of P. Viola and M. Jones's work to extract facial region from video frames.[22] However, facial videos captured in real-life scenarios can exhibit low face quality due to the problems of pose variation, varying levels of brightness, and motion blur. A low-quality face produces erroneous results in facial feature points or landmarks tracking. To solve this problem, we use an FQA module,[16,23] which calculates four scores for four quality metrics: resolution, brightness, sharpness, and out-of-plan face rotation (pose). The quality scores are compared (with thresholds[23] having values of 150×150, 0.80, 0.80, and 0.20, for resolution, brightness, sharpness, and pose, respectively) to check whether the face needs to be discarded. If a face is discarded, we concatenate the trajectory segments to remove discontinuity.[5] As we measure the average HR over a long video sequence (say, 30 to 60 seconds), discarding a few frames (fewer than 5 percent of the total frames) doesn't

greatly affect the regular characteristic of the trajectories but removes the most erroneous segments coming from low-quality face frames.

Feature Points and Landmarks Tracking

Tracking facial feature points and generating a trajectory can record head motion in facial video caused by heartbeat. Our objective with trajectory extraction and signal processing is to find the cyclic trajectories of tracked points by removing the noncyclic components from the trajectories. Because GFT-based tracking has some limitations, using voluntary head motion and facial expression changes in a video produces one of two problems: completely missing the track of feature points and erroneous tracking. We observed more than 80 percent loss of feature points by the system in such cases. In contrast, SDM doesn't miss or erroneously track landmarks in the presence of voluntary facial motions or expression change—as long as the face is qualified by FQA. Thus, the system can find enough trajectories to measure HR. However, GFT uses a larger number of facial points to track than SDM, which uses only 49 points. This causes GFT to preserve more motion information than SDM when there's no voluntary motion. We propose combining the trajectories of GFT and SDM. To generate combined trajectories, the face is passed to a GFT-based tracker to generate trajectories from facial feature points and then appended with SDM trajectories. Let the trajectories be expressed by location time-series $S_{t,n}(x, y)$, where (x, y) is the location of a tracked point n in video frame t.

Vibration Signal Extraction

The trajectories from the previous step are usually noisy due to voluntary
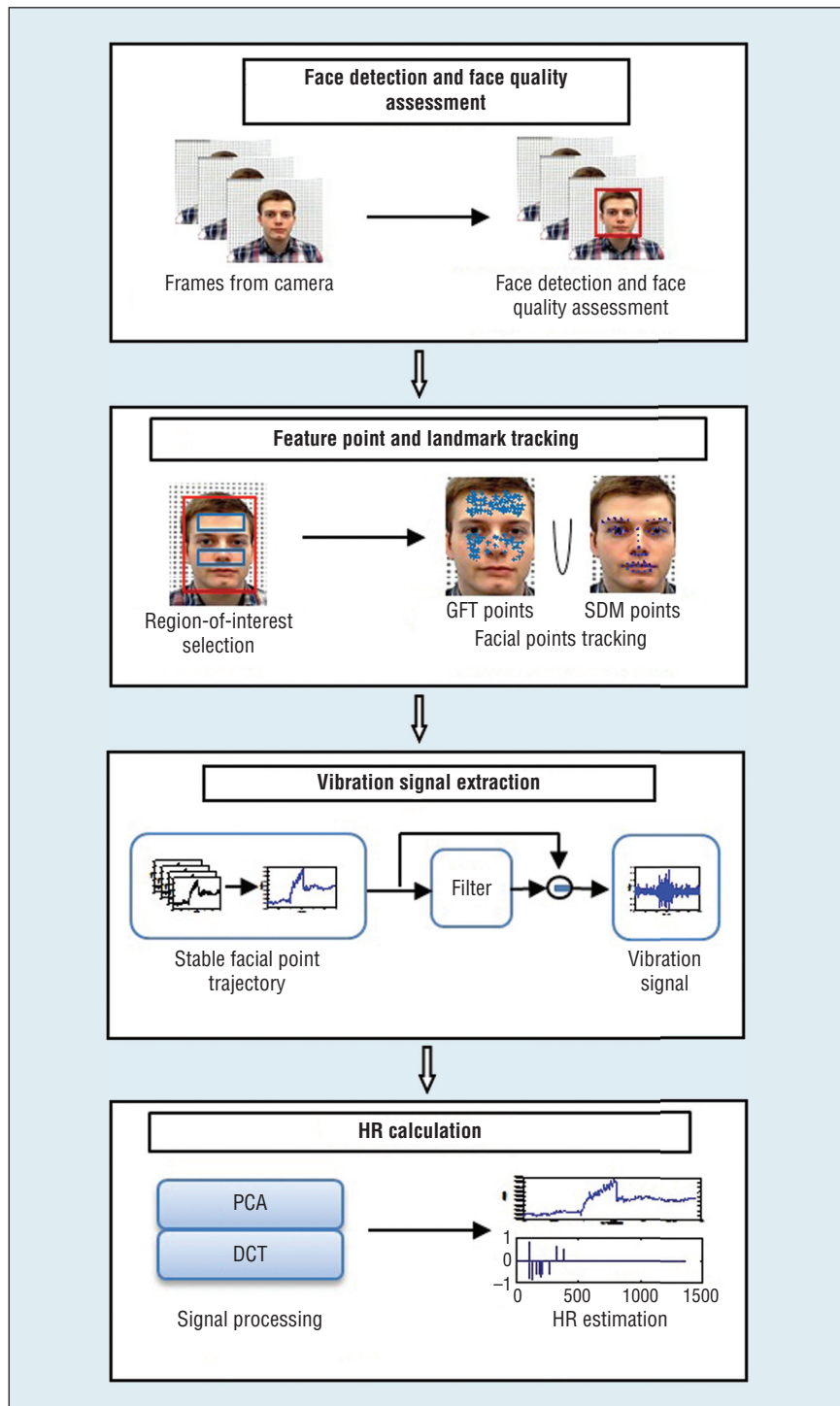
Figure 2. The block diagram of the proposed system. We acquire the facial video, track the intended facial points, extract the vibration signal associated with heartbeat, and estimate the HR.

head motion, facial expression, and/or vestibular activity. We reduce the effect of such noise by employing filters to the vertical component of each feature point's trajectories. An eighth-order Butterworth band pass filter with a cutoff frequency of 0.75 to 5.0 Hz (human HR lies within this

range[11]) is used along with a moving average filter:

$$S_n\left(t\right)=\frac{1}{w}\sum_{i=-\frac{w}{2}}^{\frac{w}{2}-1}S_n\left(t+i\right),\,\frac{w}{2}<t<T-\frac{w}{2},\tag{6}$$

where w is the length of the moving average window (length is 300 in our experiment) and T is the total number of frames in the video. These filtered trajectories are then passed to the HR measurement module.

HR Measurement

As head motions can originate from different sources and only those caused by blood circulation through the aorta reflect HR, we apply a principal component analysis (PCA) algorithm to the filtered trajectories (S) to separate the motion sources. PCA transforms S to a new coordinate system by calculating the orthogonal components P via load the matrix L as follows:

$$p = S \cdot L,\tag{7}$$

where L is a $T \times T$ matrix with columns obtained from the eigenvectors of $S^T S$. Among these components, the most periodic one belongs to heartbeat.[11] We apply discrete cosine transform (DCT) to all the components (P) to find the most periodic one.[12] We then employ fast Fourier transform (FFT) on the inverse DCT of the component and select the first harmonic to obtain HR.

Experimental Environment and Datasets

We implemented our proposed method by using a combination of Matlab (SDM) and C++ (GFT with KLT) environments. We used three databases to generate results: a local database for demonstrating FQA effects, a local database for HR measurement, and the publicly available MAHNOB-HCI database.[24] For the first database, we collected 6 datasets of 174 videos from 7 subjects to conduct an experiment to report the effectiveness of employing FQA in the proposed system. We put four webcams (Logitech C310) at 1-, 2-, 3-, and 4-meter distances to acquire facial video with four different

face resolutions of the same subject. The room's lighting condition was changed from bright to dark and vice versa for the brightness experiment. Subjects were requested to have around 60 degrees of out-of-plan pose variation for the pose experiment. The second database contained 64 video clips and defined 3 scenarios to constitute our own experimental database for HR measurement ex-periment, which consists of about 110,000 video frames of about 3,500 seconds. These datasets were captured in two different setups: an experi-mental setup in a laboratory and a real-life setup in a commercial fitness center. In scenario 1 (normal), subjects exposed their faces in front of the cameras without any facial expression or voluntary head motion (about 60 seconds). In scenario 2 (internal head motion), subjects made facial expressions (smiling/laughing, talking, and angry) in front of the cameras (about 40 seconds). In scenario 3 (external head motion), subjects made voluntary head motion in different directions in front of the cameras (about 40 seconds).

The third database (MAHNOB-HCI) has 491 sessions of videos that

Table 1. Dataset names, definitions, and sizes.

Name	Definition	No. videos
Lab_HR_Norm_Data	Video data for HR measurement collected for lab scenario 1	10
Lab_HR_Expr_Data	Video data for HR measurement collected for lab scenario 2	9
Lab_HR_Motion_Data	Video data for HR measurement collected for lab scenario 3	10
FC_HR_Norm_Data	Video data for HR measurement collected for fitness center scenario 1	9
FC_HR_Expr _Data	Video data for HR measurement collected for fitness center scenario 2	13
FC_HR_Motion_Data	Video data for HR measurement collected for fitness center scenario 3	13
MAHNOB-HCI_Data	Video data for HR measurement collected from other sources[24]	451
Res1, Res2, Res3, Res4	Video data acquired from 1-, 2-, 3-, and 4-meter distances, respectively, for the face quality assessment experiment	29×4
Bright_FQA	Video data acquired during lighting changes for FQA experiment	29
Pose_FQA	Video data acquired during pose variations for FQA experiment	29

are longer than 30 seconds and to which subjects consent via a "yes" attribute. Among these sessions, data for subjects 12 and 26 were missing. We collected the rest of the sessions as a dataset for our experiment, which are hereafter called MAHNOB-HCI_ Data. We used 30 seconds (frames 306 to 2,135) from each video for HR measurement and the corresponding ECG signal for the ground truth.[5] Table 1 summarizes all the datasets used in our experiment.

Performance Evaluation

Our proposed method uses a combination of the SDM- and GFT-based approaches for trajectory generation from the facial points. Figure 3 shows the calculated average trajectories of tracked points in two experimental videos. We included the trajectories obtained from GFT[13,18] and SDM[16,17] for facial videos with voluntary head motion, along with some example video frames depicting face motion. As is shown in the figure, GFT and SDM provide similar trajectories when there's little head motion. When the voluntary head motion is sizable, the GFT-based method fails to track the point accurately and thus produces an erroneous trajectory because of large optical flow. However, SDM provides stable trajectory in this case, as it doesn't suffer from large optical flow. We also observe that the SDM trajectories provide more sensible amplitude than the GFT trajectories, which in turn contributes to clear separation of heartbeat from the noise.

Unlike other work,[11] the proposed method utilizes a moving average filter before employing PCA on the trajectory obtained from the tracked facial points and landmarks. Figure 4a shows the effect of this moving average filter, which reduces noise, softens extreme peaks in voluntary head motion, and provides a smoother
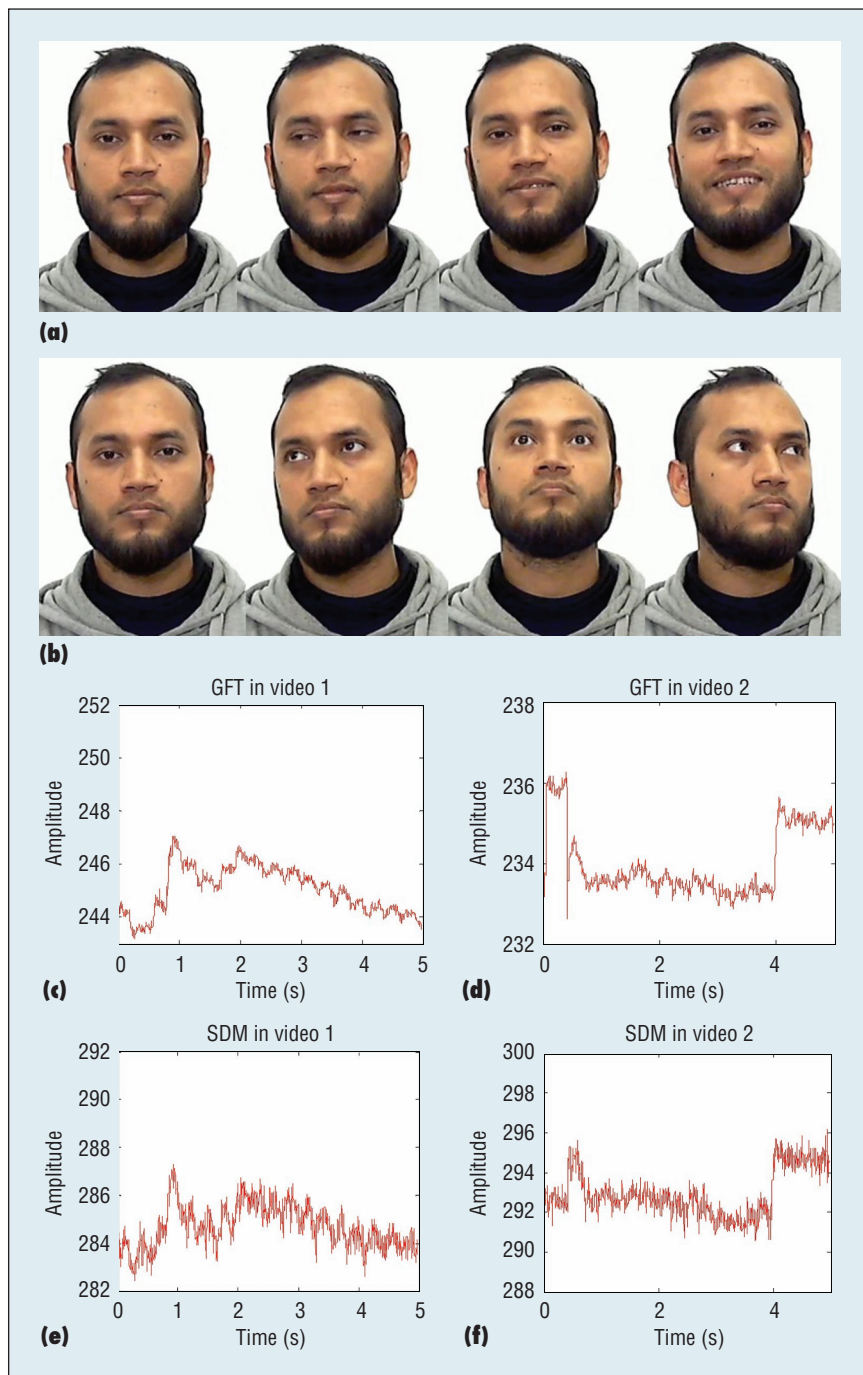

Figure 3. The effect of voluntary head motion in trajectories from 5 seconds of two experimental video sequences with small motion (video1) and large motion at the beginning and end (video2). Example frames depict (a) small and (b) large motion from a video, and trajectories of tracking points extracted by GFT[18] in (c) and (d) and SDM[17] in (e) and (f), respectively.

signal to PCA in the HR detection process.

Our proposed method uses DCT instead of FFT[11] to calculate the periodicity of the cyclic head motion signal. Figure 4b shows a trajectory of head motion from an experimental video and its FFT and DCT representations
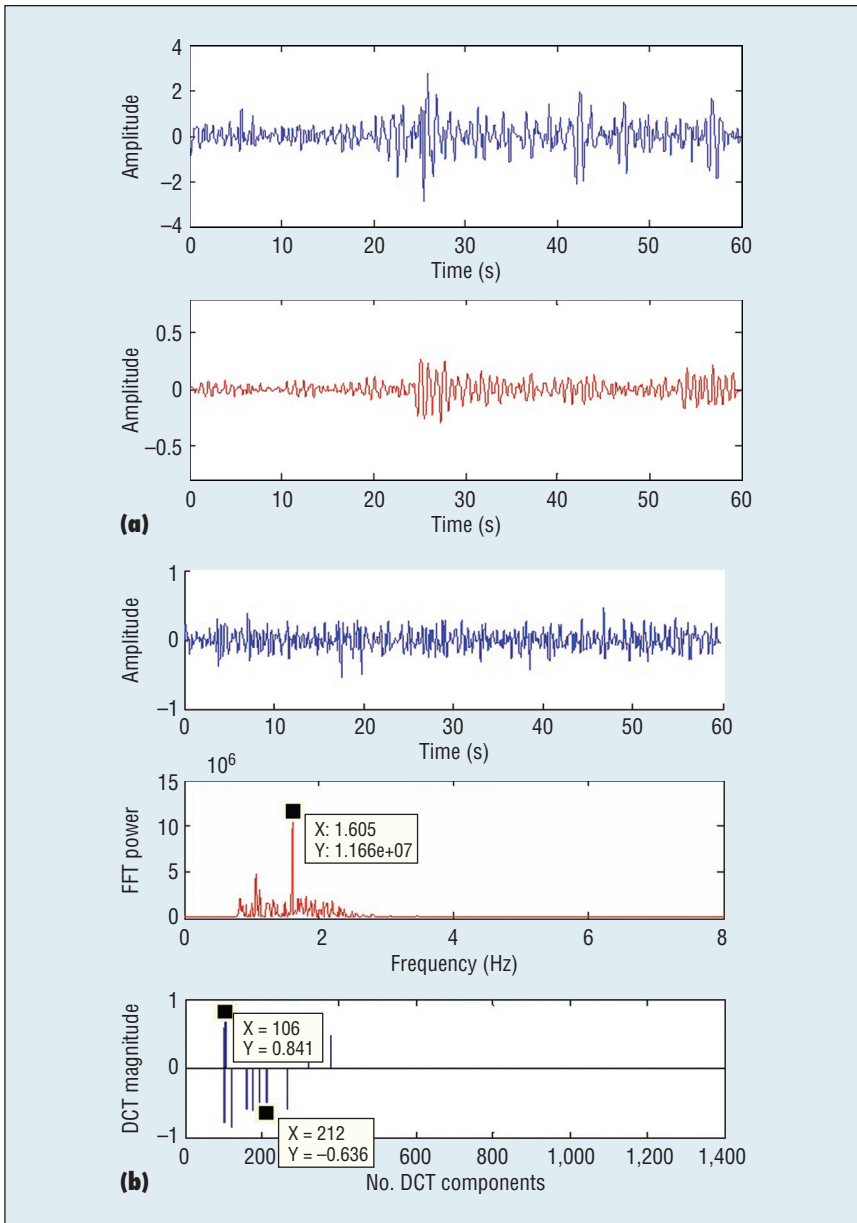
Figure 4. The effect of (a) the moving average filter on the trajectory of facial points to get a smoother signal by noise and extreme peaks reduction and (b) the difference between extracting the periodicity (heartbeat rate) of a cyclic head motion signal by using fast Fourier transform (FFT) power and discrete cosine transform (DCT) magnitude.

after preprocessing. In the figure, we see that the maximum power of FFT is at frequency bin 1.605. This, in turn, gives an HR of $1.605 \times 60 = 96.30$, whereas the actual HR obtained from the ECG was 52.04 beats per minute (bpm). Thus, the method that used FFT in the HR estimation[11] doesn't always produce good results. On the other hand, using DCT[12] yields a result of 52.35 bpm from the selected DCT component $X = 106$. This is very close to the actual HR.

We also conducted an experiment to demonstrate the effect of employing FQA in the proposed system. The experiment had three sections for three quality metrics: resolution, brightness, and out-of-plan pose. Table 2 shows the results of HR measurement on six datasets. From the results, it's clear that when resolution decreases, system accuracy decreases accordingly. Thus, FQA for face resolution is necessary to ensure a good-sized face in the system. The results also show that brightness and pose variation influence HR measurement. When frames of low quality in terms of brightness and pose are discarded, HR measurement accuracy increases.

Performance Comparison

We compared the performance of the proposed method against state-of-the-art methods[3,5,6,11,12] on the experimental datasets listed in Table 1. Table 3 lists the accuracy of the proposed method's HR measurement results compared with motion-based state-of-the-art methods[11,12] on our local database. We measured the accuracy in terms of percentage of measurement error: the lower the error generated by a method, the higher that method's accuracy. From the results, we observe that the proposed method showed consistent performance, although the data acquisition scenarios were different with different datasets. By using both GFT and SDM trajectories, the proposed method gets more trajectories to estimate the HR pattern in the case of HR_Norm_Data and accurate trajectories due to non-missing facial points in the cases of HR_Expr_Data and HR_Motion_Data. On the other hand, the previous methods suffer from fewer trajectories or erroneous trajectories from the data acquired in challenging scenarios. (One showed an up to 25.07 percent error in HR estimation from videos having facial expression change.) The proposed method outperforms the previous methods in both environments (lab and

fitness center) of data acquisition for all three scenarios.

Table 4 shows the performance comparison of HR measurement by our proposed method and state-of-the-art methods (both color- and motion-based) on MAHNOB-HCI_Data. We calculate the root mean square error (RMSE) in bpm and mean error rate in percentage to compare the results. We can observe that Li's,[5] Irani's,[12] and the proposed method showed considerably higher results than the other methods because they take into consideration the presence of voluntary head motion in the video. However, unlike Li's color-based method, Irani's method and the proposed method are motion-based. Thus, changing the illumination condition in MAHNOB-HCI_Data doesn't greatly affect the motion-based methods, as indicated by the results. Finally, we observe that the proposed method outperforms all these state-of-the-art methods in HR measurement accuracy.

Our proposed system's performance for HR measurement is highly accurate and reliable not only in a laboratory setting with no-motion, no-expression cases in artificial light in the face[11,12] but also in challenging, real-life environments. However, the proposed system isn't adapted yet to the real-time application for HR measurement due to dependency on temporal stability of the facial point trajectory. ▢

References

1. J. Klonovs et al., *Distributed Computing and Monitoring Technologies for Older Patients*, 1st ed., Springer, 2015.
2. H.-Y. Wu et al., "Eulerian Video Magnification for Revealing Subtle Changes in the World," *ACM Trans. Graphics*, vol. 31, no. 4, 2012, pp. 65:1–65:8.
3. M.-Z. Poh, D.J. McDuff, and R.W. Picard, "Non-contact, Automated Cardiac Pulse Measurements Using Video Imaging and Blind Source Separation," *Optics Express*, vol. 18, no. 10, 2010, pp. 10762–10774.
4. H. Monkaresi, R.A. Calvo, and H. Yan, "A Machine Learning Approach to Improve Contactless Heart Rate Monitoring Using a Webcam," *IEEE J. Biomedical and Health*

Table 2. Analyzing the effect of the face quality assessment in heartbeat rate (HR) measurements.

Experiment name	Dataset	Average error (%) in HR measurement
Resolution	Res1	10.65
	Res2	11.74
	Res3	18.86
	Res4	37.35
Brightness	Bright_FQA before FQA	18.77
	Bright_FQA after FQA	17.62
Pose variation	Pose_FQA before FQA	17.53
	Pose_FQA after FQA	14.01

Table 3. Performance comparison between the proposed and the state-of-the-art methods of HR measurement on local database.

Dataset name	Average error (%) in HR measurement		
	Balakrishnan and colleagues[11]	Irani and colleagues[12]	Proposed method
Lab_HR_Norm_Data	7.76	7.68	2.80
Lab_HR_Expr_Data	13.86	9.00	4.98
Lab_HR_Motion_Data	16.84	5.59	3.61
FC_HR_Norm_Data	8.07	10.75	5.11
FC_HR_Expr_Data	25.07	10.16	6.23
FC_HR_Motion_Data	23.90	15.16	7.01

Table 4. Performance comparison between the proposed and the state-of-the-art methods of HR measurement on MAHNOB-HCI database.

Method	Root mean square error (bpm)	Mean error rate (%)
Poh and colleagues (2010)[3]	25.90	25.00
Kwon and colleagues[7]	25.10	23.60
Balakrishnan and colleagues[11]	21.00	20.07
Poh and colleagues (2011)[6]	13.60	13.20
Li and colleagues[5]	7.62	6.87
Irani and colleagues[12]	5.03	6.61
The proposed method	3.85	4.65

Informatics, vol. 18, no. 4, 2014, pp. 1153–1160.

5. X. Li et al., "Remote Heart Rate Measurement from Face Videos under Realistic Situations," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2014, pp. 4321–4328.

6. M.-Z. Poh, D.J. McDuff, and R.W. Picard, "Advancements in Noncontact, Multiparameter Physiological Measurements Using a Webcam," *IEEE Trans. Biomedical Eng.*, vol. 58, no. 1, 2011, pp. 7–11.

7. S. Kwon, H. Kim, and K.S. Park, "Validation of Heart Rate Extraction Using Video Imaging on a Built-In Camera System of a Smartphone," *Proc. 2012 Ann. Int'l Conf. IEEE Eng. in Medicine and Biology Soc.*, 2012, pp. 2174–2177.

8. M.A. Haque, K. Nasrollahi, and T.B. Moeslund, "Heartbeat Signal from Facial Video for Biometric Recognition," *Image Analysis*, R.R. Paulsen and K.S. Pedersen, eds., Springer, 2015, pp. 165–174.

9. M.A. Haque, K. Nasrollahi, and T.B. Moeslund, "Can Contact-Free Measurement of Heartbeat Signal Be Used in Forensics?," *Proc. 23rd European Signal Processing Conf.*, 2015, pp. 1–5.

10. M.A. Haque, K. Nasrollahi, and T.B. Moeslund, "Efficient Contactless Heartbeat Rate Measurement for Health Monitoring," *Int'l J. Integrated Care*, vol. 15, no. 7, 2015, pp. 1–2.

11. G. Balakrishnan, F. Durand, and J. Guttag, "Detecting Pulse from Head Motions in Video," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2013, pp. 3430–3437.

12. R. Irani, K. Nasrollahi, and T.B. Moeslund, "Improved Pulse Detection from Head Motions Using DCT," *Proc. 9th Int'l Conf. Computer Vision Theory and Applications*, 2014, pp. 1–8.

13. J. Bouguet, "Pyramidal Implementation of the Lucas Kanade Feature Tracker," tech. report, Intel Microprocessor Research Labs, 2000.

14. A.D. Bagdanov et al., "Posterity Logging of Face Imagery for Video Surveillance," *IEEE Multimedia*, vol. 19, no. 4, 2012, pp. 48–59.

15. K. Nasrollahi and T.B. Moeslund, "Extracting a Good Quality Frontal Face Image From a Low-Resolution Video Sequence," *IEEE Trans. Circuits and Systems for Video Tech.*, vol. 21, no. 10, 2011, pp. 1353–1362.

16. M.A. Haque, K. Nasrollahi, and T.B. Moeslund, "Quality-Aware Estimation of Facial Landmarks in Video Sequences," *Proc. IEEE Winter Conf. Applications of Computer Vision*, 2015, pp. 1–8.

17. X. Xiong and F. De la Torre, "Supervised Descent Method and Its Applications to Face Alignment," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2013, pp. 532–539.

18. J. Shi and C. Tomasi, "Good Features to Track," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 1994, pp. 593–600.

19. T.F. Cootes, G.J. Edwards, and C.J. Taylor, "Active Appearance Models," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 23, no. 6, 2001, pp. 681–685.

20. A.U. Batur and M.H. Hayes, "Adaptive Active Appearance Models," *IEEE Trans. Image Processing*, vol. 14, no. 11, 2005, pp. 1707–1721.

21. Y. Feng et al., "Exploiting Temporal Stability and Low-Rank Structure for Motion Capture Data Refinement," *Information Science*, vol. 277, 2014, pp. 777–793.

22. P. Viola and M.J. Jones, "Robust Real-Time Face Detection," *Int'l J. Computer Vision*, vol. 57, no. 2, 2004, pp. 137–154.

23. M.A. Haque, K. Nasrollahi, and T.B. Moeslund, "Real-Time Acquisition of High Quality Face Sequences from an Active Pan-Tilt-Zoom Camera," *Proc. 10th IEEE Int'l Conf. Advanced Video and Signal Based Surveillance*, 2013, pp. 443–448.

24. M. Soleymani et al., "A Multimodal Database for Affect Recognition and Implicit Tagging," *IEEE Trans. Affective Computing*, vol. 3, no. 1, 2012, pp. 42–55.

THE AUTHORS

Mohammad A. Haque is a PhD fellow in the Visual Analysis of People Lab, Aalborg University. His research interests include vision-based patient monitoring, biometrics, and decision support systems. Haque received an MS in computer science and information technology from University of Ulsan, South Korea. Contact him at mah@create.aau.dk.

Ramin Irani is a PhD fellow in the Visual Analysis of People Lab, Aalborg University. His research interests include facial expression recognition, social cue recognition, and soft biometrics. Irani received an MS in electrical engineering from Blekinge Institute of Technology, Sweden. Contact him at ri@create.aau.dk.

Kamal Nasrollahi is an associate professor in the Visual Analysis of People Lab, Aalborg University. His research interests include facial analysis systems, biometrics recognition, soft biometrics, and inverse problems. Nasrollahi received a PhD in computer vision from Aalborg University, Denmark. Contact him at kn@create.aau.dk.

Thomas B. Moeslund is the head of the Visual Analysis of People Lab, Aalborg University. His research focuses on all aspects of the automatic analysis of images and video data. Moeslund received a PhD in computer vision from Aalborg University, Denmark. Contact him at tbm@create.aau.dk.

cn *Selected CS articles and columns are also available for free at http://ComputingNow.computer.org.*