Remote Liveness and Heart Rate Detection from Video

Yunbin Deng
Yunbin.deng@baesystems.com

1/10/2021

BAE SYSTEMS

Outlines

- Motivations
- Design of experiments
- Video-based liveness detection algorithms
- Video-based heart rate estimation algorithms
- Experimental results
- Conclusions and Future work

Motivations

- The remote detection of liveness is critical for disaster response, the military, and law enforcement
- Existing solutions are mostly based on special sensor hardware or the spectral signature of living skin

- This paper uses commercial electro-optical and infrared (EO/IR) sensors to capture a very short video
    - Stationary and moving subject
    - Stationary and moving sensor platform
    - EO and IR sensor modality

- The key advantages are:
    - low cost
    - requires very short video
    - works with many parts of a human body even when skin is not visible
    - works on any motion caused by eyes, mouth, heartbeat, breathing, or body parts
    - works in all lighting conditions

Design of Experiments

Environment	Indoor		Outdoor	
Sensors	Logitech Webcam	FLIR ONE PRO	Camera on DJI drone	SONY 4K camcorder
Frame rate	24	8.7	23	29
Range (m)	1, 3	1	21	50/100/150
# of subject	3	2	3	4
Purpose	LD	LD	LD	LD, HRD
Experiment	III.B,C,D	III.B	III.A,F	III.E,IV.A

Video-based Remote Liveness Detection System

The liveness detection system works via the following steps:

1. Detect a person and/or face. This generates a bounding box for each region of interest (ROI).
2. For each ROI, use the bounding box to crop the images of all subsequent frames.
3. calculate the mean square error (MSE) between subsequent sub-image and the first sub-image.
4. Estimate the sensor/environmental noise level. Take the MSE between the second and first frame.
5. Adaptive sensor noise cancellation. Subtract the sensor noise from all the MSEs between all subsequent frames and the first frame.
6. Apply a threshold to make the liveness decision and/or
7. Plot a heat map of micro-motion for human in the loop detection.

Low Resolution Person and Face Detection

- Depending on the standoff range, the number of pixels on a person or face can vary widely.
- A multi-scale deep neural network (DNN) is trained on a multi-scale human body and face dataset
- The DNN is then used to perform a multi-scale search for a person and face
- Allows robust detection of a person or a partial view of a person when the body image is less than 50 pixels tall

Figure 1. A video frame from a DJI hovering drone shows three people on the ground. A DNN tiny face detection algorithm detected all three low resolution faces.

Facial Micro-motion Detection from EO/IR Videos

Let *X* represents a facial Region of Interest (ROI) sub-image, which is a matrix containing *r* rows and *c* columns of pixels. This can be a facial area for face-based micro-motion detection. Let *X(n), n=1,2,…,M*, represent *M* consecutive ROI matrixes. We compute the MSE between subsequent ROI to the first ROI and get a new MSE sequences *Y(k), k=1,2,…, M-1.*

$$Y(k) = MSE(X(k+1), X(1)), \text{ for } k=1,2,.., M-1,$$

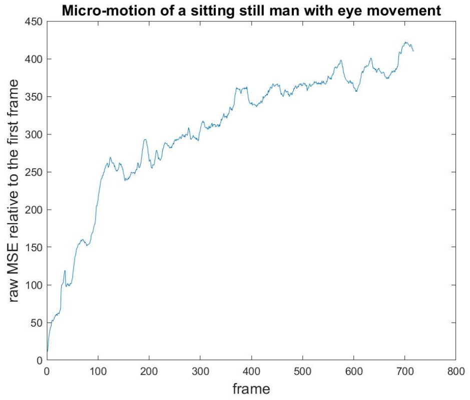

Figure 4. A webcam MSE time series based on a detected facial region of a man sitting still with some eye movement.

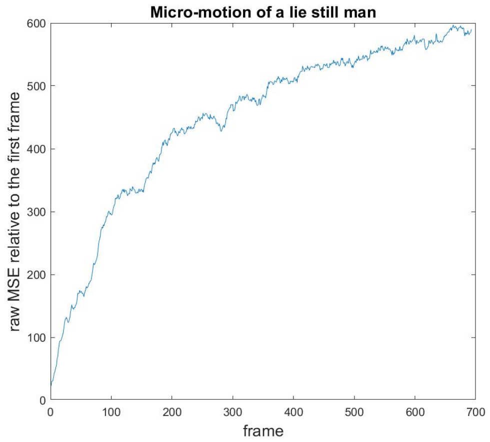

Figure 5. A webcam MSE time series based on a detected facial region of a man lying still sideway.

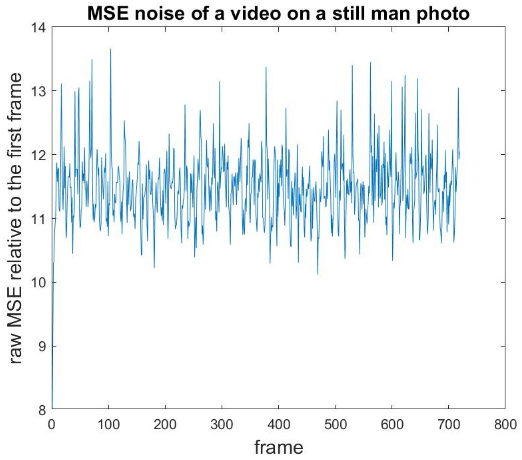

Figure 7. MSE time series based on a detected facial region of a still photo. It only show sensor noise signal.

Environmental and Sensor Noise Compensation

Sensor noise normalized MSE as *Y_norm(k)*, which can be derived from the raw MSE time series *Y(K)* as follows:

$$Y\_norm(k) = Y(k+1) - Y(0), \quad for\ k=1,2,..,M-2,$$

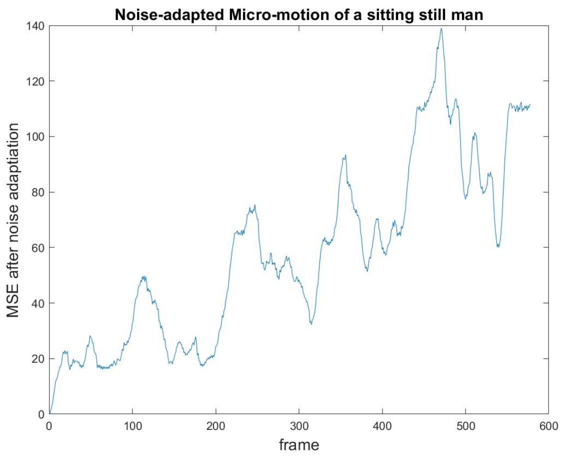

Figure 8. Sensor noise compensated MSE time series based on a detected facial region of a man sitting still.

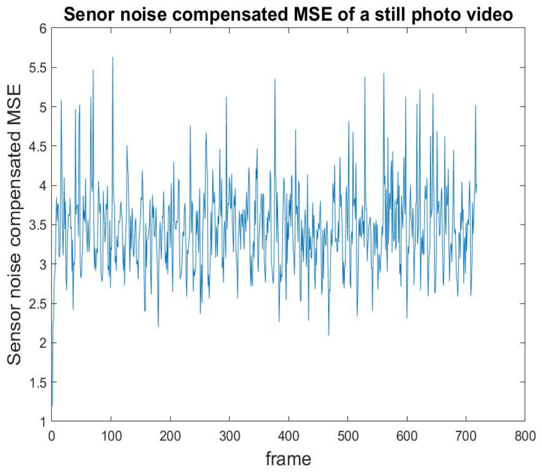

Figure 9. Sensor noise compensated MSE time series based on a detected facial region of a still photo video. It shows only lighting noise signal.

Effectiveness of micro-motion detection from other parts of human body

The motion heat map clear show the magnitude of micro-motion corresponding to the chest and right foot area.

Figure 10. The full body video of a person lying still and flat on a carpeted floor. Only the edge from the first frame of the video is shown.

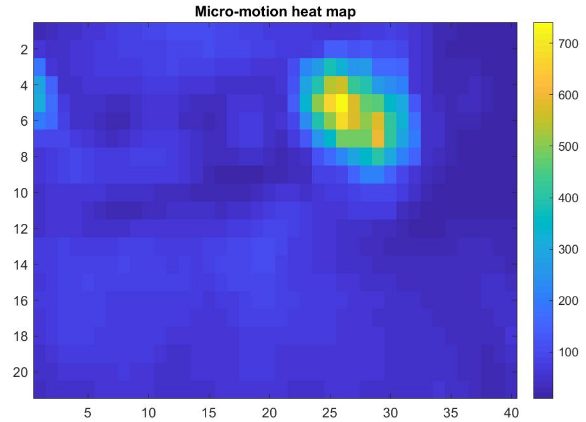

Figure 11. The maximum of MSE time series of windowed grids on the video, shown as a heat map of micro-motions of the video in Figure 10**.**

The Impact of Longer Standoff Range and Lower Resolution Video

Simulate the impact of greater distance and less image resolution, the original video frames are down sampled by a factor of 10.

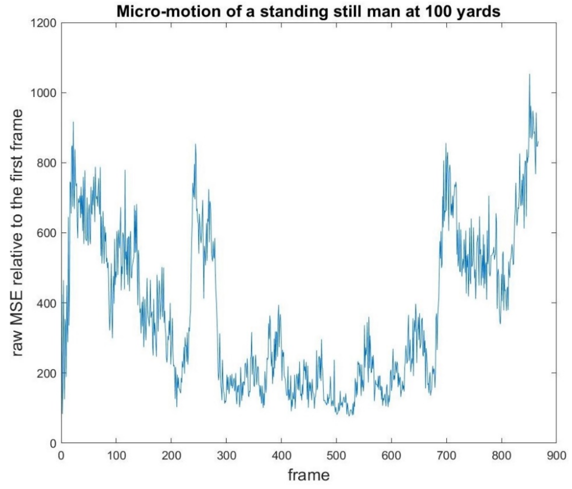

Figure 12. MSE time series based on a detected facial region of a man standing still at 100 meters range.

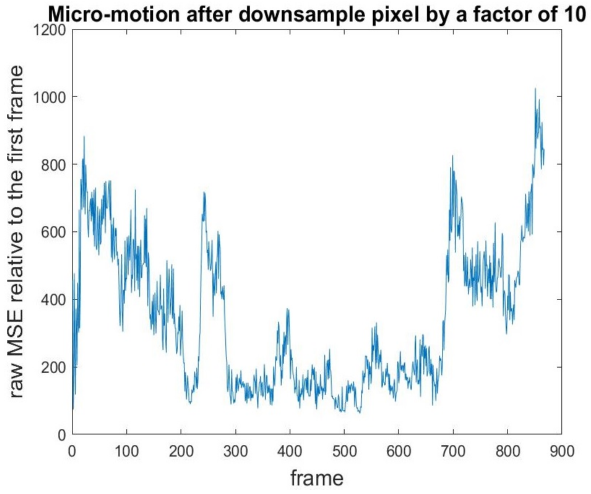

Figure 13. Same plot as Figure 12, but the original video resolution was down sampled by a factor of 10. The impact on MSE is very small

Sensor Platform Motion Compensation

The video stabilization algorithm works via the following steps:

1. Apply fast corner detection algorithm to generated salient points in each image frame.
2. For each two consecutive image frames, find the corresponding feature points using Fast Retina Key point (FREAK) descriptor and Hamming distance [19].
3. Estimate a robust geometric affine transform from these noisy point correspondences using a variant of the RANSAC algorithm, called M-estimator Sample Consensus (MSAC) algorithm [20].
4. Compute an accumulative scale-rotation-translation transform based on each pair of local transforms to represent the platform motion since the first frame.

The video stabilization algorithm achieves real-time computation with a frame rate of 12 FPS on a 2.7GHz CPU laptop

Figure 1. Video frame from a drone

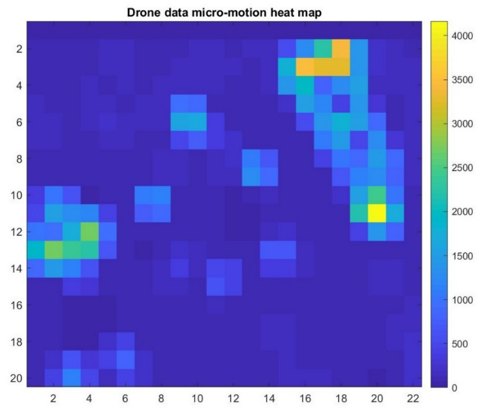

Figure 14. Zoom-in view of micro-motion heat map for video frame shown in Figure 1. There are three persons lying still on grass. The heat map captured micro-motion of heads, hands, and legs of three person.

Remote Outdoor Heart Rate Detection from Video

- Remote photoplethysmography (rPPG): video-based standoff heart rate estimation
- Blood circulation causes skin color change
- Most existing work are based on short standoff range of a few feet
- Key challenges are noise lighting changes introduced by environment and motion

Our rPPG algorithm consists of the following steps:

1. Apply Dlib for 68 face landmarks detection [13], applied once for every 1 second of data.
2. The facial region, excluding mouth, eyes, eyebrow, and nose, is divided into 5 by 5 pixels grids.
3. The average pixel value of green channel is used to represent the signal of interest for each grid.
4. A 1-d time series is extracted for each grid. The number of grid is fixed, assume the subject is stationary.
5. A singular value decomposition (SVD) is applied on the high dimensional spatiotemporal time series.
6. A heuristic signal quality metric is defined to select a subset of high quality components in the SVD [15].
7. Short time Fourier transform (STFT) is then applied to constructed signal.
8. The highest energy frequency component of STFT is the estimated heart rate at each time step, or
9. Perform dynamic programming to search the 'optimum' smoothed heart rate estimation curve

100 Meters Heart Rate Estimation Experimental Results

- we use video data recording at 100 meters range on four adult subjects standing still.
- A SONY Exmor R 4k camcorder is used for the video recording.
- The ground truth is read from the oximeter at the end of each video recording.

- Each video is about 30 seconds long
- STFT is updated every 5 seconds.
- We take the mean of all STFTs as the estimate.

Table 2: 100 Meters Standoff rPPG Performance

Subject	rPPG HR	Oximeter HR	rPPG Error
1	84	90	6.6%
2	74.4	93	20%
3	80.6	94	14.3%
4	75	79.5	5.7%
Average	78.5	89.1	11.7%

- Reasonable good accuracy at 100 meters
- The performance varies among subjects.

rPPG Performance Benchmark on a Public Domain Dataset

The COHFACE data
- 160 one-minute-long videos from 40 subjects, with 12 females and 28 males
- The test set contains 64 videos.
- Videos are recorded with a standard webcam and data are sampled at 20 PFS.
- The data considers the case of well controlled lighting condition and natural indoor lighting conditions.

To extract the 'ground truth' heart rate time series, we designed a peaking detection based method to the Blood-Volume Pulse data.

Our algorithm achieved average Root Mean Square Error (RMSE) of 8.67 beat per minute (BPM) on the test set of COHFACE.

Our algorithm outperforms all evaluated algorithms in published work on the COHFACE data

Table 3: rPPG Algorithm performance comparison on public available COHFACE indoor dataset

rPPG Algorithm	RMSE (BPM)
2SR [27]	25.84
CHROM [25]	12.45
LiCVPR [26]	25.59
HR-CNN [18]	10.78
Ours without smoothing	9.55
Ours	8.34

Applications

- *Law enforcement and military application*
  - The capability of detecting life or death at a safe distance can thus save lives.

- *Disaster response*
  - Using unmanned ground vehicles or unmanned aircraft systems, with a liveness detection enabled sensor, can conduct a more effective search for survivors.

- *Commercial and health care*
  - Health monitor during pandemic. Standoff detection reduces risk of virus spreading
  - Baby and child monitoring for daycare and home.
  - Senior care at nursing homes
  - Hospital, prison, depressed people with high risks of suicide
  - Airport, border, and other security check point, the standoff detection of abnormal heart rate enables quick identification of potential high-risk suspects warranting further investigation.

Conclusions and Future work

Conclusions:

- Proof of concept of video-based liveness detection on stationary and UAV platform
- Standoff heart beat detection at outdoor 100-meter range

Future work

- Liveness Detection
    - Developing a fully functional UGV and UAV system works in a cluttered environment.
    - Improve computational speed, system robustness, and performance on field test

- Heart Rate Estimation
    - Consistent performance across subjects
    - Robustness to free body and head movement while performing daily routines