Emotion & Heartbeat Detection using Image Processing

Kaushal Kanakia, Saurabh Patil, Sujay Sabnis, Vedant Shah

Abstract— A non-contact means of measuring heart rate and detecting emotions could be beneficial for sensitive populations, and the ability to calculate pulse using a simple webcam or phone camera could be useful in telemedicine.Emotions are one important aspect of human life. Emotions define us, allow us to express ourselves, and react to the outside world's events. They are fundamental to the human experience, influencing our everyday tasks such as learning, communicating and rational decision-making.Happiness, anger, excitement, surprise, among other emotions motivate certain actions and enrich human experience. Heart rate, measured in beats per minute (BPM), can be used as an index of an individual's physiological state. Each time the heart beats, blood is expelled and travels through the body. This blood flow can be detected in the face using a standard webcam that is able to pick up subtle changes in color that cannot be seen by the naked eye. Due to the light absorption spectrum of blood, we are able to detect differences in the amount of light absorbed by the blood traveling just below the skin (i.e., photoplethysmography).Emotions can be detected by using haar cascades, feature extraction from ROI and then processing it.OpenCV and Python has been used for human computer interaction.For the detection of heartbeat, we use remote photoplethmysography.This paper discusses the reimplementation of one such approach that uses independent component analysis on mean pixel color values within a region of interest (ROI) about the face.

Index Terms— Remote Photo Plethmysography, ROI (Regon of Interest), ICA (Independent Component Analysis), PCA (Principal Component Analysis) OpenCV, Python.

1 Introduction

A person's heart rate can be indicative of their health, fitness, activity level, stress, and much more. Cardiac pulse is typically measured in clinical settings using electro cardio-gram (ECG), which requires patients to wear chest straps with adhesive gel patches that can be abrasive and become uncomfortable for the user. Heart rate may also be monitored using pulse oximetry sensors that may be worn on the fingertip or earlobe. These sensors are not convenient for long-term wear and the pressure can become uncomfortable over time. In addition to the discomforts of traditional pulse measurement devices, these devices can damage the fragile skin of premature new-borns or elderly people. For these populations especially, a non-contact means of detecting pulse could be very beneficial. Non-contact heart rate measurement through a simple webcam or phone camera would also aid telemedicine and allow the average person to track their heart rate without purchasing special equipment.

Through a computer or phone camera may help detect changes in a per-son's heart rate over time and indicate changing health or fitness. Heart rate can be detected without contact through photo-plethysmography (PPG), which measures variations in blood volume by detecting changes in light reflectance or transmission throughout the cardiovascular pulse cycle. PPG is usually performed with dedicated light sources with red or infrared wavelengths, as is the case for pulse oxime-try sensors.Verkruysse showed that the plethysmographic signal could also be detected in video from a regular colour camera. They found that the signal could be detected within the red, green, and blue channels of colour video of exposed skin, but that it was strongest in the green channel, which corresponds to the fact that haemoglobin has absorption peaks for green and yellow light wavelengths. They also found that although the signal could be detected in multiple locations on the body, it was strongest on the face, especially on the forehead. Although the plethysmographic signal may be detected in the raw colour channel data, it is mixed in with other sources of colour variation such as changes in ambient light.Emotions play an important role in detection of emotions.When a person is happy,his heart rate is normal but when the person is excited or nervous,his heart rate increases.Thus we can use the various emotions to detect the changes in his heartbeat.

2 Technical Approach

Detecting Heartrate and Emotions consists of 3 sections which are mentioned below.Firstly the facial region must be detection.Secondly, desired region of interest (ROI) must be chosen within face bounding box. And third, the plethysmographic signal and emotions must be extracted from the change in pixel colours within the ROI over time and analysed to determine the prominent frequency within the heart rate range.

2.1 Face Detection and Tracking

Face detection and tracking is performed using Haar cascade classifiers as proposed by Viola and Jones and improved by Lienhart et al. [6]. Specifically, we use the OpenCV Cascade Classifier pre-trained on positive and negative frontal face images. The face detector is built from a cascade of classifiers of increasing complexity, where each classifier uses one or more Haar-like features.The features consist of two, three, or four rectangular pixels as shown in figure 1. Each feature is calculated as the sum of pixels in the grey rectangles less the sum of pixels in the white rectangles. These features are able to detect simple vertical, horizontal, and diagonal edges and blobs.Since there are over 180,000 potential features in each sub-window, only a small subset of these features are actually used. The AdaBoost learning algorithm is used to train classifiers built on one to a few hundred features. To choose which feature(s) to use, a weak classifier is trained on each feature

individually and the classification error is evaluated. The classifier (and associated feature) with the lowest error is chosen for that round, the weights are updated, and the process is repeated until the desired number of features are chosen. This process creates a single strong classifier that is a weighted combination of numerous weak classifiers. The strong classifiers are then used in series in the attentional cascade, which is essentially a decision tree for each sub-window within the image.

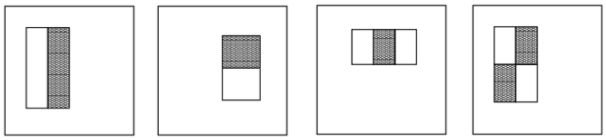

Figure 1. Example Haar-like features used in the boosted cascade face classifier from the original paper by Viola and Jones [11]. Each feature is calculated as the sum of pixels in the grey rectangles less the sum of pixels in the white rectangles.

2.2 Region of Interest (ROI)

Since the face bounding box found using face detection contains background pixels in addition to the facial pix-els, an ROI must be chosen from within the bounding box. The simplest choice of ROI is to use the center 60% of the bounding box width and the full height, as was done by Poh et al. [8]. Since the bounding box is usually within the face region height-wise but outside the face width-wise, this method simply adjusts the box to exclude background pixels to the sides of the face. With this method, some hair or background pixels are usually still present at the corners of the box.

We also explore other means of selecting the ROI. We examine the effects of removing the eye region, which contains non-skin pixels that may vary across frames due to blinking or eye movement. Removing pixels from between 25% and 50% of the bounding box height worked well to remove the eyes. We also explore retaining only the pixels above the eye region, since Verkruysse et al. found the fore-head has the strongest plethysmographic signal.

Facial segmentation is performed using GrabCut, developed by Rother et al. GrabCut segments images by iteratively minimizing an energy cost function. This energy minimization can be achieved by defining a graph model to represent the image and determining the minimum cut for the graph to yield two sets of nodes, which represent the foreground and background

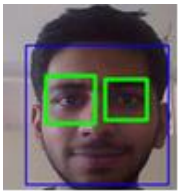

Figure 2:Haar Cascade to select features

2.3 Heart Rate Detection

Using Facial Landmarks, we detect the emotions in real timing by comparing it with the conoha database.Once we have an ROI for each frame, we can begin to extract the heart rate from the colour image data. The first step is to average the pixels in the ROI across each colour channel to get three signals xR(t), xG(t), and xB(t) corresponding to the average red, green, and blue facial pixels at time t. We then normalize these signals across a 30-second sliding window with a 1-second stride (so the heart rate is re-estimated every second).

We then use ICA to extract the independent source signals from the observed mixed colour signals. ICA assumes that the number of source signals is no more than the number of observed signals, so we assume there are three source signals s1(t), s2(t), and s3(t) contributing to the observed colour changes in the three channels. ICA assumes the observed mixed signals are a linear combination of these source signals. Although this assumption may not be valid as changes in blood volume and the intensity of reflected light in skin tissue over distance may be nonlinear, for the 30-second time window it should be a reasonable approximation.. Then ICA attempts to find an approximation that maximizes the non-Gaussianity of each source. We can use Fast ICA from the scikit-learn library to recover the approximate source signals s(t) .Once we have the source signals, we can apply a Fourier transform to the data to examine their power spectrum and determine the prominent signal frequencies. We can iso-late frequency peaks in the power spectrum within the range 0.75 to 4 Hz, which corresponds to physiological heart rate ranges of 45 to 240 bpm. The measured heart rate will be the frequency within the acceptable range corresponding to the peak with the highest magnitude.

3 Experimental Setup

The person was placed in front of the webcam of the computer and the face bounding box was applied to the face on the real time video.The highest peak within the power spectrum of the signal was taken to be the reference heart rate for each 30-second window.

In addition to determining the robustness of the algorithm to subject movement, we also examine the robustness to bounding-box noise. Although the bounding box was usually well centered on the face in each frame, we could imagine that in a noisier environment in which there is more movement of either the camera or subject, poorer lighting, facial occlusions, or more background clutter, there could be more error in the location of the facial bounding box. To simulate this, we can add artificial noise to the bounding box corner locations found with the Haar cascade classifier. In each frame, corners of the bounding box were shifted horizontally and vertically by percentages of the width and height, where the percentages were chosen uniformly at random up to a maximum noise percent.

4 RESULTS

Once the facial bounding box (es) found using the Haar cascade classifiers were narrowed down to a single bounding box on the subject's face. Results of various ROI selection options, including a nar-rower box, a box with the eyes removed, a box around the forehead, or a segmented face, may be seen in figure 4. The simplest ROI, namely the narrower bounding box that was used by Poh et al., usually contained mostly skin pixels for frontal images, reimplementation of GrabCut. As can be seen, the segmentation excludes most hair and background pixels within a couple iterations.

The next step of the algorithm is to find the heart rate from the selected ROI pixels. As described in the technical approach, this process includes finding the mean RGB pixel values within the ROI for each frame and then normalizing across a 30-second window, ICA to extract independent source signals, and power spectrum analysis to determine the prominent frequencies.

although there was also sometimes hair or background appearing at the corners. When the subject tilted or turned their head, more background pixels become part of the ROI, as shown in figure 2.

Figure 3 shows the original bounding box, adjusted bounding box, and segmentation results for a straight face using the

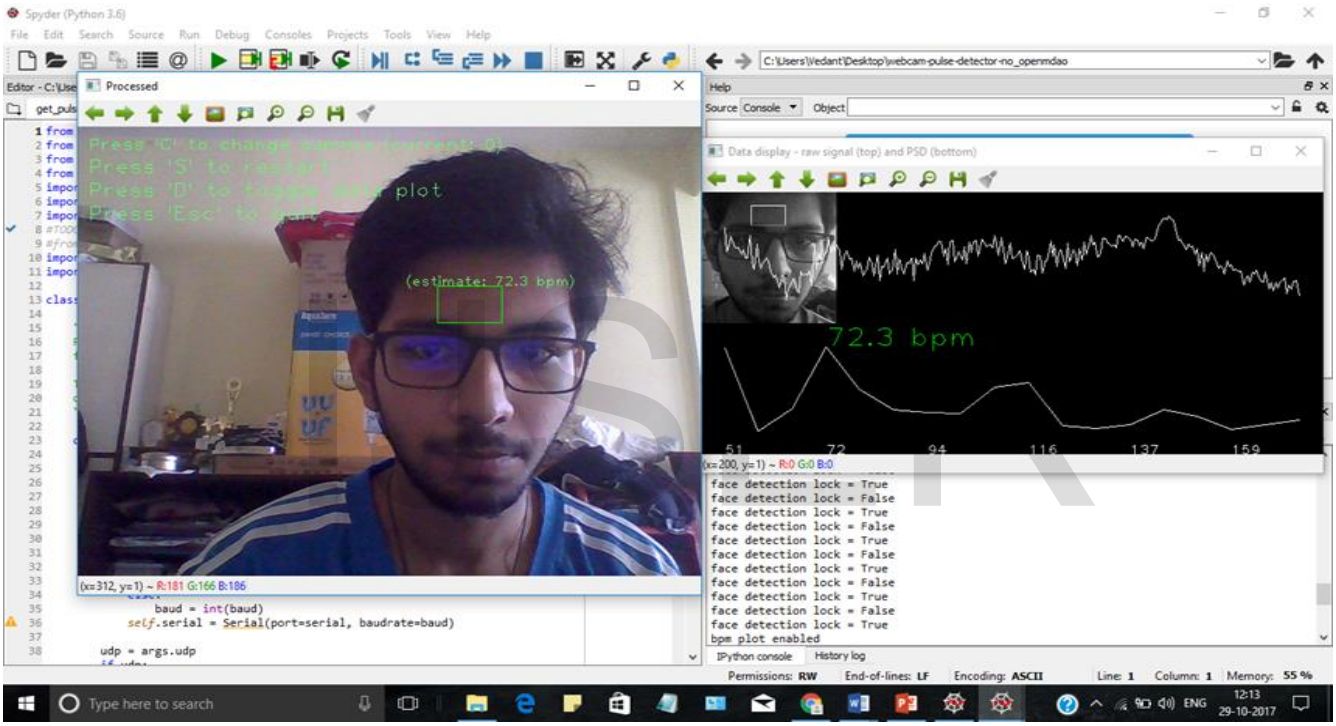

Figure 3: Detecting the heartrate by locking on to the fore head movement

5 Conclusion

We have seen that heart rate may be measured in regular colour video of a person's face. We observed heart rate errors of 3.4 - 0.6 bpm for videos of still faces and 2.0 - 1.6 bpm for videos with movement. Since the calculated heart rate was consistently lower than the reference in all videos and with small standard deviation, it is possible that the base error is due to a miscalculation in video frame rate or finger pulse sensor sample rate. If the Samsung cam-era had a frame rate that was 5% higher than the reported frame rate, or if the Arduino sample rate was actually 5% lower than desired, or if there were a smaller error in both rates, this would account for the observed heart rate calculation errors. Future studies could use a medical-grade pulse monitor to ensure that the reference heart rate is as accurate as possible. If it is determined that there is truly always a consistent offset between the reference and the calculated heart rate as observed in this study, a calibration step could be used to remove the bias. Such a system could potentially measure heart rate in still video to within 0.6 bpm.

Although sample rate errors could cause a consistent bias in heart rate measurements, the standard deviation and percentage of outlier measurements provide a better indication of the robustness of the algorithm to changes in the video. We observed under 5% outlier measurements for still faces and under 20% for videos of moving faces. Since these are both well under 50%, outliers could be easily identified and removed from heart rate calculations even if there were no reference heart rate to use as a basis. We also found that as bounding box noise increased, choosing an ROI by segmenting out facial pixels helped to diminish the noise effects and keep the outliers low.

Overall we find that for clean videos of subjects' faces in good lighting, using a simple ROI defined by 60% of the width and the full height of the facial bounding box works just as well as removing the eye region or segmenting out facial pixels, and outperforms a box only around the fore-head region. Since this simple box ROI is much faster than performing GrabCut to segment out the facial pixels (more than 12 times as fast), the simple box would be the suggested method for such videos performing GrabCut to segment out the facial pixels (more than 12 times as fast), the simple box would be the suggested method for such videos.

Another future study could attempt to do video heart rate calculations in real-time rather than with pre-recorded videos. Currently, using a simple box ROI, a one-minute video may be processed on a MacBook Pro (2.6 GHz) in under two minutes, processing about 8 fps. This may in fact already be fast enough for real-time if a camera with a frame rate under 8 fps were used to record the videos. Since the maximum physiological heart rate expected by the algorithm is 4 Hz, the sample rate would be just enough to avoid aliasing at 8 fps (8 Hz is the Nyquist frequency). Since 4 Hz is a highly unlikely heart rate especially for a person at rest, the frame rate could likely be decreased even further and the algorithm max heart rate adjusted accordingly. If segmentation is required (for example if there is significant bounding box noise), other adjustments would be required to speed up the algorithm processing time.per, do not replicate the abstract as the conclusion. A conclusion might elaborate on the importance of the work or suggest applications and extensions. Authors are strongly encouraged not to call out multiple figures or tables in the conclusion—these should be referenced in the body of the paper.

References

[1] G. Balakrishnan, F. Durand, and J. Guttag. Detecting pulse from head motions in video. CVPR, 2013.

[2] R. Bhatt. Embedded lab.

[3] Y. Boykov and M.-P. Jolly. Interactive graph cuts for optimal boundary and region segmentation of objects in n-d images. International Conference on Computer Vision, 2001.

[4] Y. Boykov and V. Kolmogorov. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. IEEE Transactions of PAMI, 26(9):1124–1137, 2004.

[5] G. Bradski. The OpenCV Library. Dr. Dobb's Journal of Software Tools, 2000.

[6] R. Leinhart, A. Kuranov, and V. Pisarevsky. Empirical ana-lyis of detection cascades of boosted classifiers for rapid ob-ject detection. MRL Technical Report, 2002.

[7] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Ma-chine learning in Python. Journal of Machine Learning Re-search, 12:2825–2830, 2011.

[8] M.-Z. Poh, D. McDuff, and R. Picard. Non-contact, auto-mated cardiac pulse measurements using video imaging and blind source separation. Optical Society of America, 2010.

[9] C. Rother, V. Kolmogorov, and A. Blake. "grabcut" – inter-active foreground extraction using iterated graph cuts. SIG-GRAPH, 2004.

[10] W. Verkruysse, L. Svaasand, and J. S. Nelson. Remote plethysmographic imaging using ambient light. Opt Express, 2008.

[11] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. Accepted Conference on Com-puter Vision and Pattern Recognition, 2001.