Measurement of Heart Rate from Video Capturing

Jingyi Li, Theo Braune
Institut Polytechnique de Paris - École Polytechnique
theo.braune@polytechnique.edu, jingyi.li@ip-paris.fr

December 14, 2021

Abstract

In the following we present a video-based heart rate measurement method, which could be of great benefit in the field of telemedicine or surveillance. To do this, in the first step we will make use of the Grab-Cut method to extract the information from the color values in the skin of the face. We describe how we apply the Independent Component Analysis (ICA) in order to obtain a signal that can be further analyzed with the Fourier Transformation for the sake of extracting the frequency spectrum of color values in the skin, from which we will deduce the heart rate.

Keywords: Computer Vision, Signal Processing

1 Introduction

Heart rate is a physical parameter that is of big importance for human health. It is an indicator of the level of physical stress. Conventional methods of measuring heart rate often use contact-based methods such as a chest strap, wristband or electrodes. In practice, however, it is often a problem to measure heart rate in this way over a longer period of time. People are therefore interested in measuring heart rate without contact. With the advent of telemedicine, there is also a need for this, as it is otherwise difficult in practice. There is also an interest from fitness apps that measure physical parameters and activity over a longer period of time to get this important value without additional devices. In his keynote talk in "New Ways of Thinking of the Mobile Phone for Healthcare" at the Heidelberg Laureate Forum 2019, Shwetak Patel describes on a more general level how the use of computer vision on the human body can be used to perform certain routine medical checks even in areas with poor medical care [Pat19]. Another possible field of application for such a method could be the use in combination with video surveillance at security-relevant locations such as airports, where authorities are interested in recognising possible endangerers with abnormal vital values and, if necessary, taking timely action. In the following, we will present a method that makes it possible to determine the heart rate from a frontal video recording of a face.

The biological basis for this theory of so-called photo-plethysmography is presented, for example, in [All07]. Allen et al, show that there is a biological connection between the colour values in the skin and the current blood volume, so that the change in the colour values can give an indication of the change in the blood volume and thus the heart rate. The improvement in the resolution of modern cameras makes it possible to measure these changes. In [VSN08] it is shown for the first time how this can work with a camera in ambient light.

In the following, we will first use GrabCut as a facial segmentation method to extract the skin of the face. In principle, [All07] shows that one can also use any area of the skin, but the changes in the face are best measurable. We will take the average of the colour values on the segmented face for each frame in order to carry out the signal processing over a time window.

In practice, these colour signals are often disturbed by changes in ambient light or movement. To extract non-Gaussian signals, we use the Independent Component Analysis (ICA), which was presented the first time in [JH91]. Following this, we will determine the frequency with the help of Fourier transformation and thus measure the heart rate.

In section 2 we will present the implemented method in detail. In section 3 we will present the results obtained and then discuss limitations of our method.

1.1 Related Work

The biological basis for our method was primarily developed in [All07] and [VSN08]. The technical implementation of these biological observations, which we followed most closely in our implementation, are outlined in [PMP10] and [Bus16]. In both methods, colour signals were extracted in segmented face areas and then the frequency spectrum was determined using a Fourier transform. Another method for contactless heart rate measurement is presented in [BDG13]. Balakrishnan et al show that it is possible to determine fixed points on the forehead whose movement pattern can then be used to determine the rate of change of blood volume in the forehead. A similar approach is taken by [RS20], who use a radar method to determine the fixed points on the face.

1.2 Experimental Setup

For our project we used a basic build-in webcam in a laptop (Tuxedo InfinityBook pro14) with a resolution of 1,0 MP. The Videos were all recorded in color and with a frame-rate of 30

FPS. The videos were recorded with a resolution of 640x480. To obtain a ground truth for our algorithm we were using the heart rate sensor included in an iPhone and additionally we were measuring our heart rate manually by pressing a finger on our wrist and using a stop watch. We tested the algorithm under different lighting conditions, with artificial light, sunlight and different intensities.

2 Method

Measuring the heart rate from video capturing consists of three main steps. The first step is detecting the face region from the image. The second step is choosing the Region of Interest (ROI) from the face region. The third step is extracting heart rate from the sequence of ROI.

2.1 Face detection

Face detection is performed with Haar cascode classifiers which was proposed by Viola and Jones[VJ01] and improved by Lienhart et al[LM02]. It is a machine learning based approach where a cascade function is pre-trained on images with faces and images without faces. To be specific, we use the OpenCV Cascade Classifier [Bra00] in our code to detect faces for our own images.

In order to distinguish images with faces and without faces, we need to extract features from images. Haar features are what we used and they are similar to convolutional kernel. Each feature is a single value obtained by subtracting sum of pixels under the white rectangle from sum of pixels under the black rectangle.

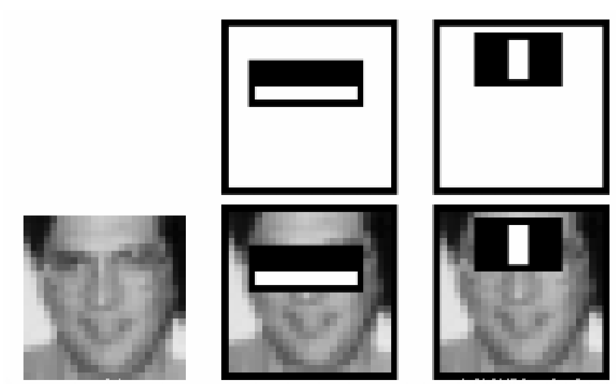

Figure 1: Illustration on how to extract Haar features from an image

For one image, we take each 24×24 windows and apply 6000 features to the image.

Then the features are grouped into different stages of classifiers and applied one-by-one. If a window fails the first stage, we discard it and don't consider the remaining features on it. If it passes, apply the second stage of features and continue the process. The window which passes all stages is a face region.

2.2 Determination of the Region of Interest

The face detection algorithm divides the images into foreground and background with a bounding box and the ROI must be chosen within it. The first choice is the forehead. We choose pixels between $\frac{1}{20}$ and $\frac{1}{6}$ of the bounding box height and between $\frac{1}{4}$ and $\frac{3}{4}$ of the bounding box width, which works well.

Another choice is the face segmentation from GrabCut developed by Rother et al[RKB04]. GrabCut is an image segmentation method based on graph cuts. With a bounding box obtained from our face detection algorithm around the object, the algorithm estimates the color distribution of the target object and that of the background using a Gaussian mixture model. Then a graph is built from this pixel distribution. Nodes in the graphs are pixels and additional two nodes are added, source node and sink node. Every foreground pixel is connected to source node and every background pixel is connected to sink node. The weights of edges connecting pixels to source node or end node are defined by the probability of a pixel being foreground or background. The weights between the pixels are defined by the edge information or pixel similarity. If there is a large difference in pixel color, the edge between them will get a low weight.

Then a min-cut algorithm is used to segment the graph. It cuts the graph into two separating source node and sink node with minimum cost function. The cost function is the sum of all weights of the edges that are cut. After the cut, all the pixels connected to Source node become foreground and those connected to Sink node become background. The process is continued until the classification converges. For our purpose we will use the implemented Grabcut Algorithm in the OpenCV Library [Bra00].

2.3 Heart rate detection

After taking a ROI for each frame, we calculate the average values for RGB color channels of the ROI, denoted as $x_R(t), x_G(t), x_B(t)$ where t is the time. Then we use independent component analysis (ICA) to extract the source signals from the observed mixed color signals.

The ICA is defined as follows: The data is represented by the observed random vector $\boldsymbol{x} = (x_1, \ldots, x_m)^T$ and the hidden components as the random vector $\boldsymbol{s} = (s_1, \ldots, s_n)^T$. The task is to transform the observed data $\boldsymbol{x}$, using a linear static transformation $\boldsymbol{W}$ as $\boldsymbol{s} = \boldsymbol{W}\boldsymbol{x}$, into a vector of maximally independent components $\boldsymbol{s}$ measured by some function $F(s_1, \ldots, s_n)$ of independence.

We use ICA by assuming that the source color signals are potentially non-Gaussian signals and that they are statistically independent from each other. Although this assumption might not be true if we consider the changes of blood volume and light intensity, but for a time window less then one minute, it is a reasonable approximation. The transformation in the ICA is linear, so the number of source signals is no more than the number of observed signals, and we assume there are three source signals $s_R(t), s_G(t), s_B(t)$ contributing to the observed color changes in the three channels. According to the definition of ICA, we have

$$s(t) = W^{-1}x(t)$$

where

$$x(t) = [x_R(t), x_G(t), x_B(t)]^T, s(t) = [s_1(t), s_2(t), s_3(t)]^T$$

and W is a 3×3 linear transformation (matrix).

We used an implemented FastICA function in the scikit-learn library [PVG$^+$11] to recover the approximate source signals $s(t)$.

After extraction of the source signals $s(t)$, we can transform the signals on time into signals on temporal frequency using Fourier Transformation. The measured heart rate will be the frequency with highest magnitude in the range from 0.75 to 3 Hz, which corresponds to physiological heart rate ranges from 45 to 180 bpm.

3 Results

We experimented with a time window of 5-15 seconds over which we then computed the Fourier transformation to obtain the frequency. When using a frame rate of 30 FPS, we are buffering 150-450 frames to our rate computing method. We realized that a smaller frame rate also yields passable results, but we observed a higher stability if augmenting the frame rate. We first only used an empirical ROI, where we cut a forehead out of the face, to obtain the first results, that matched our ground truth, but we observed a high instability when the head is moved. The classifier with the GrabCut segmentation keeps some stability. When the head is only moving in the horizontal plane, we observed good results, and for some cases like a video call this is a reasonable assumption.

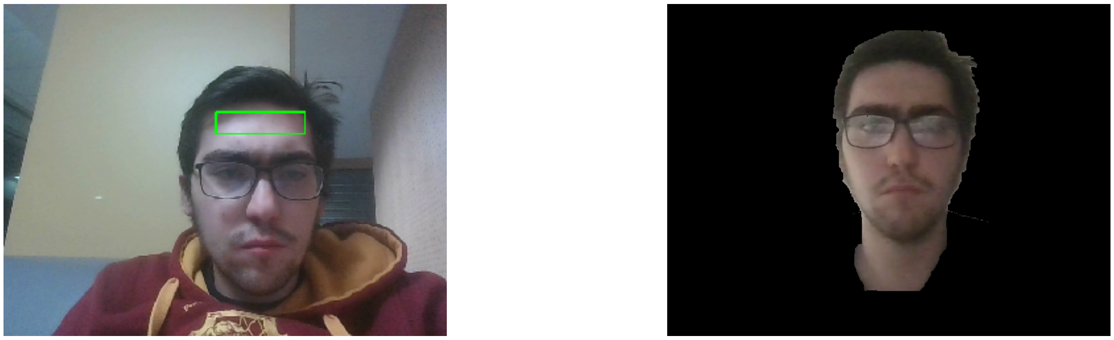

Figure 2: Illustration of our two different methods to extract a region of interest for the signal processing. Left is the result of primitive forehead detection, and right is the result of GrabCut segmentation.

We observed that in our experimental setup, the GrabCut is in fact the bottleneck for computation time. For a 640x480 picture, one step takes around a second. Thus for a low resolution it is in theory possible to run the code with the GrabCut in real time, but our method works better on a recorded video, where we compute the face segmentation frame by frame as a pre-processing operation. If we extract the data from the ICA and transform the signal into frequency space, we fetch the frequency peak in the range of 0.75 and 3Hz. This value will be converted to bpm and corresponds to the heart rate.

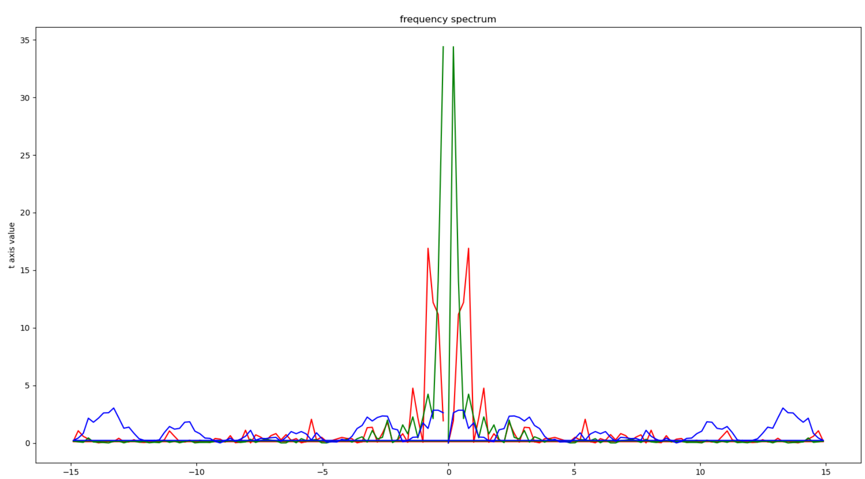

Figure 3: Example of the frequency spectrum.

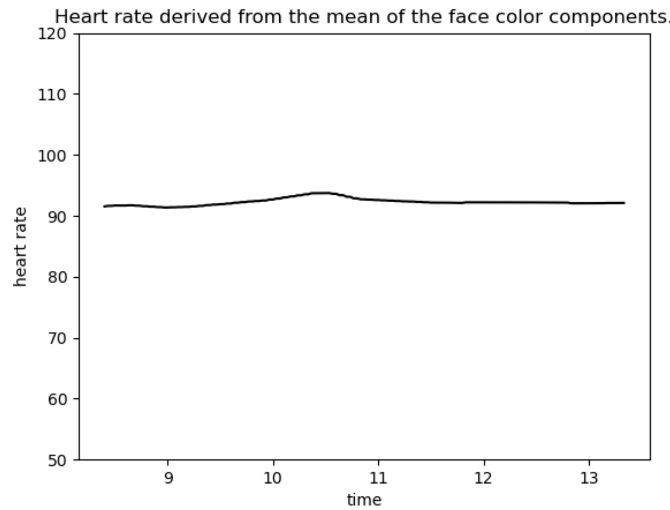

Figure 4: Example of the plot of the measured heart rate.

In our tests, we obtained highest accuracy, when we used the mean of the three color frequency peaks. In the described setup, the frequency is measured with an error in the range of ± 5 bpm from the ground truth.

In Figure 4 we used a 5 second time window to compute the mean over all heart rate values. The plotted value at 15 seconds actually corresponds to the the mean of all heart rates over the time window. This is necessary, as the raw heart rate data is noisy. Furthermore it is reasonable to assume, that the heart rate stays constant over a 5-15 second time window. In the first five seconds, the plot will show the noisy data. After five seconds we will use the mean value over the time window; the plot will look more stable.

4 Failure Cases and Difficulties

We have detected instability in our method due to external illumination. Empirically it turned out that the algorithm works the best when there is no sunlight and there is uniform artificial light. For us it was a surprise that in our tests, the derived heart rate from the mean of the frequency peak of the three color channels performed best in our tests, although Verkruysse et al [VSN08] give a biological explanation, that the green color channel should perform best. We could not find a satisfying explanation whether this is due to the lack of quality of our code or for some other reason.

We discovered that Make-Up on the face decreases the performance of our method, as it might

Figure 5: Example of a lighting situation that causes difficulties.

cover the changes of color in the skin. This is why we had difficulties in applying our method to interview or movie scenes.

We made a similar observation for people with a darker skin color.
We observed a certain stability when we move the head, although the moving can be a problem if different parts of the face are visible over times.

Figure 6: Make-Up in interview scenes can effect the performance of our method.

5 Conclusion and further ideas

We have seen that our method provides a beginning step for the contactless heart rate measurement.
Nevertheless there are still big downsides for our method in practise. It would be highly problematic if applying our method to a large part of the population with darker skin color.

A possible solution for this issue is drawn out in [BDG13], where Balakrishnan et al do not use the color change of the skin, but use fixed points on the skin and observe the movement of them. Then -similar to our method- they use these signals to derive the frequency of the change of blood volume using the Fourier Transformation. Also it is possible to use Convolutional Neural Networks in order to improve the face segmentation quality [NMT+18].
To conclude this work we want to emphasize the reader to think and discuss about the ethical implications that these technology could have. Already with our method it is possible to collect personal sensitive data like heart rate in a video conference even if the person does not notice it. A possible starting point for this discussion is given in [Noo20] and we emphasize the reader who might continue working on this to keep these considerations in mind.

The code for this project can be found under `https://github.com/JeansLli/X-INF573-heart-rater`.

References

[All07] John Allen. Photoplethysmography and its application in clinical physiological measurement. *Physiological Measurement*, 28:R1–39, 04 2007.

[BDG13] Guha Balakrishnan, Fredo Durand, and John Guttag. Detecting pulse from head motions in video. pages 3430–3437, 2013.

[Bra00] G. Bradski. The OpenCV Library. *Dr. Dobb's Journal of Software Tools*, 2000.

[Bus16] Isabel Bush. Measuring heart rate from video, 2016.

[JH91] Christian Jutten and Jeanny Herault. Blind separation of sources, part i: An adaptive algorithm based on neuromimetic architecture. *Signal Processing*, 24(1):1–10, 1991.

[LM02] R. Lienhart and J. Maydt. An extended set of haar-like features for rapid object detection. In *Proceedings. International Conference on Image Processing*, volume 1, pages I–I, 2002.

[NMT+18] Yuval Nirkin, Iacopo Masi, Anh Tuan Tran, Tal Hassner, Medioni, and Gérard Medioni. On face segmentation, face swapping, and face perception. In *IEEE Conference on Automatic Face and Gesture Recognition*, 2018.

[Noo20] R. Van Noorden. The ethical questions that haunt facial-recognition research. *Nature*, pages 354–358, Nov 2020.

[Pat19] Shwetak Patel. New ways of thinkingof the mobile phone for healthcare, 2019.

[PMP10] Ming-Zher Poh, Daniel J. McDuff, and Rosalind W. Picard. Non-contact, automated cardiac pulse measurements using video imaging and blind source separation. *Opt. Express*, 18(10):10762–10774, May 2010.

[PVG+11] 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: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[RKB04] Carsten Rother, Vladimir Kolmogorov, and Andrew Blake. "grabcut": Interactive foreground extraction using iterated graph cuts. New York, NY, USA, 2004. Association for Computing Machinery.

[RS20] Ciattaglia G. De Santis A. Gambi E. Ricciuti, M. and L Senigagliesi. Contactless heart rate measurements using rgb-camera and radar, 2020.

[VJ01] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. In *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, volume 1, pages I–I, 2001.

[VSN08] Wim Verkruysse, Lars Othar Svaasand, and J Stuart Nelson. Remote plethysmographic imaging using ambient light. *Optics express*, 16 26:21434–45, 2008.