CIS II Project Paper Evaluation: "A temporal video-processing method to improve heart rate estimation"

Azwad Sabik

March 11, 2015

Team 7 - Mobile Perfusion Analysis

Project Overview

Team 7, the Mobile Perfusion Analysis team, is attempting to develop an integrated software-and-hardware system that is capable of characterizing local perfusion of chronic wounds. Laser doppler imaging currently holds as the standard for accurate chronic wound perfusion assessment, but is costly and time-inefficient. Under the joint mentorship of Dr. Emad Boctor and Joshua Budman, CEO of mobile image-based wound assesment group Tissue Analytics, we hope to reach a solution that is portable, effective, and cost- and time-efficient.

Current Approach

As suggested by Budman, our team is currently investigating the efficacy of a video-processing technique known as Eulerian Video Magnification (EVM) in the extraction of perfusion-related information from skin video data acquired by mobile phone cameras. In our preliminary research on the technique and its implementation, we came across an example application of pulse extraction. While the paper we were evaluating described this example as achievable, it did not provide us algorithmic details or results from a structured experiment with the method to evaluate its effectiveness in this particular application. Perceiving pulse extraction from skin videos as being a step towards our goal of mapping over a particular region of skin, we sought out a paper that might evaluate this technique. Through our search, we came across Kohan and Nasrabadi's paper, the subject of this evaluation.

Paper

Goal

The Kohan paper appeared to be well aligned with our search criteria. It held the goals of evaluating the efficacy of an EVM-based algorithm for estimating patient technique and comparing three methods of extracting pulse-related information from the output of EVM-processed input videos.

Overview of Methods

The paper described an experiment in which 32 participants (6 males and 32 females in the age range of 10-75 years) were studied for 15 minutes each in order to provide data for analysis and validation of an EVM-based algorithm for heart rate extraction. Each participant was asked to sit still at a table in a room lit by ambient sunlight while being videoed by two webcams capable of recording video with a resolution of 640x480 pixels. One webcam was placed 400 cm from the participant's face, while the other was placed 50 cm from the participant's left hand. A photoplethysmogram (PPG) signal was additionally acquired from the participant's left hand's third finger. The experimenters aimed to use the webcam videos as input signals for their algorithm for estimating participant heart rates, and subsequently compare these signals with the heart rates indicated by the PPG. In order to increase heart rate variations throughout the experiment, each patient was exposed to five minutes of "relaxing" music, five minutes of "fast" music, and what the experimenters described as five minutes of "anything" (though it should be noted, the "anything" was played before the other two in the given order).

Once video data had been collected, the face data and the hand data were processed through two similar but distinct pipelines. Both sets of video data were divided up into ten-second segments, providing 90 data points for each of the 32 participants. The hand video data points were processed directly by the EVM algorithm with the full Laplacian pyramid and an ideal 12 Hz bandpass filter (for temporal filtering). The face video data points went through similar EVM processing, but were additionally pre-processed such that the input to the EVM algorithm was the contents of the middle 60% of the bounding box for the face (detected via MATLAB OpenCV) in each frame of the video. Once each data point had been EVM procesed, heart rate was extracted from each video by each of three methods: fast-Fourier transform (FFT), zero crossing, and peak detection. While it is unclear as to what exactly was treated as the time-series for each of these methods, we may imagine it to be a time-varying color intensity ("color series") for purpose of understanding how these methods are carried out. The authors simply took the most prominent frequency shown in the FFT of the color-series to be the heart rate for the first method; extrapolated a frequency from the distances between some average value of color intensity for the second method; and extrapolated a frequency from the distances between the peak values of color intensity for the last method. The authors ultimately compared these methods' resultant frequencies against the PPG-estimated heart rates in order to judge their effectiveness.

Summary of Findings

The paper concluded that their optimal heart rate extractions were 92.15% accurate with respect to PPG readings (CAND index, or complement of absolute normalized difference: $1 - \frac{|estimate - groundtruth|}{groundtruth}$), and that they had a mean bias (estimated heart rate - PPG heart rate) of -0.04 beats per minute (bpm). They estimated that 95% of the optimally extracted heart rates would fall within 3.22 bpm of a PPG reading, and that there was a correlation of 0.89 between the PPG heart rate and their estimated heart rate.

It is worthwhile to note that their optimal extractions were derived from videos of the participants' faces, and that these extractions were produced by peak detection based analyses of EVM magnified signals. While the CAND index of the peak detection-based face analyses was 0.9215, they were only 0.8280 and 0.8289 for the FFT-based method and zero-based method respectively. Additionally, 95% of the other two methods varied by up to 9.26 bpm and 12.78

bpm from the ground truth.

Critique

On the whole, the paper was chosen for its ability to provide our team with some well needed insight into how EVM-based blood flow related experiments might be carried out. The paper detailed what we perceived to be a very well thought out experimental protocol. It utilized a wide variety of participants, and in turn, was able to test its algorithm upon a wide range of true heart rates. It generated a multiple analytic results from each collected data point, and provided a detailed statistical analysis of its results. All of this being said, there were numerous points throughout their explanation of the heart rate estimation where more detail would have been appreciated. Note of the parameters used for the EVM algorithm was almost necessary, considering that their achievements have significantly different meaning to the paper's readers if the parameters were held constant between analyses of data points versus if they were not held constant - yet, there was no mention of the parameters used. Additionally, it was unclear as to how they dealt with the possibility of the face detection algorithm returning differently sized bounding boxes throughout the face video processing pipeline - if the bounding boxes varied in size between frames of the same video sequence, then it would be impossible to apply EVM to the data points. While there are a few methods of solving this problem, the paper did not elaborate upon which particular method it used. Finally, the paper simply didn't mention how it moved from the multidimensional EVM-magnified videos to a scalar heart rate outputs. Failing to mention all of these algorithmic details, the paper makes it extremely difficult for peers to attempt any reproduction of its results.

Impact & Relevance

General

Overall, the paper successfully validates offline heart rate extraction via the EVM processing of segmented 10-second videos of faces and hands. It identifies peak detection as being more effective than either FFT or zero crossing techniques in translating EVM-magnified videos into heart rates.

Team 7

Given our goals of extracting blood flow related information from EVM magnified videos, the paper validates that our approach is worth looking into and gives us a basis for comparison. In analyzing hand signals as being almost as effective as face signals (CAND 0.9213 as opposed to 0.9215) in estimating pulse, it suggests that we can hope to find results on skin surfaces apart from the face - this is something that was not as clear from other papers. It provides us the suggestion of segmenting video prior to EVM analysis, which may serve useful in reducing overall required computation and reducing motion artifacts from disturbing our analysis. It suggests that we ought to move forward with peak detection if we wish to accurately determine pulse at any local section of skin in our videos (assuming that pulse read at a particular location will correlate with blood flow at that location). It also suggests that we might find good results in videos illuminated by natural lighting (though it does not suggest anything about unnatural lighting producing bad results), and it provides us an idea for varying heart rate during analysis without inducing motion in the input videos.