Video Heart Rate Detection

Alex Kleiner
Computer Science Department
Binghamton University

Dan Rabinowitz
Computer Science Department
Binghamton University

I. Abstract

At the center of every person's health is their heart beat. Without it organs are not supplied with oxygenated blood and in a matter of seconds death settles in. A stressed heart beat or a slower heart beat can be fatal and even slight changes in resting heart rates can improve a person's health. Therefore, measuring a person's pulse is key in everything from professional medicine to the morning run. Currently there only exists intrusive methods to take a person's pulse: watches, finger clips, or a holter monitor (electric hook ups used in hospitals). However, with the aide of eulerian video magnification our algorithm can determine a person's pulse simply from a video of someone's face or hand. This can lead to non-intrusive methods to measuring pulse. A simple smart-phone camera could now keep track of one of your vital signs. From there the smart phone could have a number of apps that utilize this and help you maintain your health. This paper examines how our algorithm works, the results of our algorithm as well as how the eulerian video magnification can help our algorithm give better results.

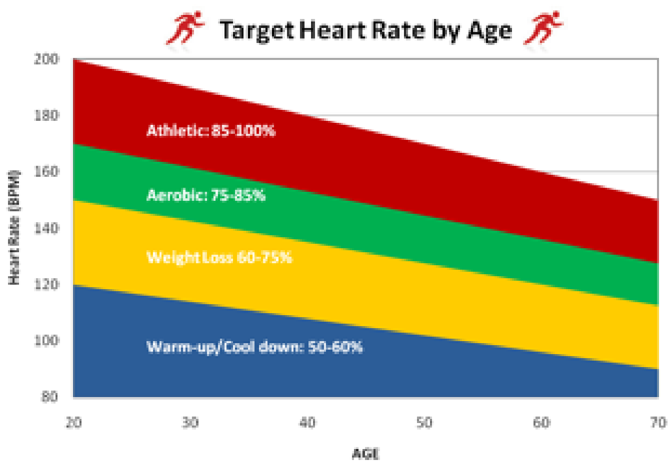
II. Introduction

The detection and measurement of human heart rate is regarded as one of the most fundamental data quantifications in the medical field. A significant number of prognoses and predictions about an individual's health can be determined from the information that heart rate can provide. Heart rate data collection is useful in a number of situations ranging from short term evaluation to long term and live detection. Heart rate data collection can be evaluated in two ways, a live sinus rhythm of each beat and an averaged sum of beats over a specified time range such as a minute. Short term detection is useful in scenarios involving athletics and for diagnosing immediate medical conditions. For an athlete or an individual participating in physical conditioning, it is common practice to evaluate one's performance and level of physical improvement based on measuring the change in sedentary and active heart rates. It is desirable for an individual to obtain a low resting heart rate between 50 and 70 beats per minute (bpm) and an active heart rate between 100 and 150 bpm. For more information on heart rate beats per minute ranges and what the data collection sets represent, see Figure 1.1 for sedentary heart rates, and for target active heart rates see Figure 1.2. When diagnosing immediate medical conditions, a medical (Figure 1.1) professional can accurately determine a variety of conditions based on a patients heart rate bpm falling out of acceptable ranges. If a patient's sinus rhythm is off and each beat of the heart happens at inconsistent intervals, a number of predictions can be made about the individuals health. When a patient has an inconsistent heart rate, this is known as arrhythmia. In long term detection, heart rate is considered the primary vital sign that a patient is still alive and stable. It is frequently used for long term situations such as in hospitals where patients are in possible serious conditions need to be monitored 24 hours a day for vital life stability by a machine without direct human observation. Easy vital sign detection is becoming increasingly more important in the present day for a large variety of reasons. Due to the significant advancements in modern medical technology over the past century, the aggregate of the United States population living at an advanced age of 65 has been increasing at a very significant rate. It is estimated that by 2030, there will be over four million United States citizens over the age of 85. This has created a significantly high demand for new methods of easy, unobtrusive, and personal health care and monitoring technology. Currently, heart rate detection technology faces a few significant caveats. The largest of these limitations is that current technology requires direct physical contact to the patient in Figure 1.2 order to achieve successful readings for data collection. The physical contact consists of electrodes on Electrocardiogram (EKG) machines, or pressure sensors on major arteries such as a wrist band. These sensors are often obtrusive and can an individual's limit range of motion and scope of freedom to explore a space. These sensors can also be uncomfortable, expensive, and produce waste when discarding used parts such as electrode stickers. Our study strives to eliminate all of the aforementioned limitations of heart rate detection technology. With each beat of the human heart, skin pigmentation changes. Skin turns flush as

	Age					
Heartrate	18-25	26-35	36-45	46-55	56-65	66+
49	Athlete	Athlete	Athlete	Athlete	Athlete	Athlete
50	Athlete	Athlete	Athlete	Athlete	Athlete	Athlete
51	Athlete	Athlete	Athlete	Athlete	Athlete	Athlete
52	Athlete	Athlete	Athlete	Athlete	Athlete	Athlete
53	Athlete	Athlete	Athlete	Athlete	Athlete	Athlete
54	Athlete	Athlete	Athlete	Athlete	Athlete	Athlete
55	Athlete	Excellent	Athlete	Athlete	Athlete	Athlete
56	Excellent	Excellent	Athlete	Athlete	Athlete	Excellent
57	Excellent	Excellent	Excellent	Athlete	Excellent	Excellent
58	Excellent	Excellent	Excellent	Excellent	Excellent	Excellent
59	Excellent	Excellent	Excellent	Excellent	Excellent	Excellent
60	Excellent	Excellent	Excellent	Excellent	Excellent	Excellent
61	Excellent	Excellent	Excellent	Excellent	Excellent	Excellent
62	Good	Good	Excellent	Excellent	Good	Good
63	Good	Good	Good	Excellent	Good	Good
64	Good	Good	Good	Good	Good	Good
65	Good	Good	Good	Good	Good	Good
66	Above Average	Above Average	Good	Good	Good	Above Average
67	Above Average	Above Average	Above Average	Good	Good	Above Average
68	Above Average	Above Average	Above Average	Above Average	Above Average	Above Average
69	Above Average	Above Average	Above Average	Above Average	Above Average	Above Average
70	Average	Above Average	Above Average	Above Average	Above Average	Average
71	Average	Average	Average	Above Average	Above Average	Average
72	Average	Average	Average	Average	Average	Average
73	Average	Average	Average	Average	Average	Average
74	Below Average	Average	Average	Average	Average	Below Average
75	Below Average	Below Average	Average	Average	Average	Below Average
76	Below Average	Below Average	Below Average	Average	Below Average	Below Average
77	Below Average	Below Average	Below Average	Below Average	Below Average	Below Average
78	Below Average	Below Average	Below Average	Below Average	Below Average	Below Average
79	Below Average	Below Average	Below Average	Below Average	Below Average	Below Average
80	Below Average	Below Average	Below Average	Below Average	Below Average	Poor
81	Below Average	Below Average	Below Average	Below Average	Below Average	Poor
82	Poor	Poor	Below Average	Below Average	Poor	Poor
83	Poor	Poor	Poor	Below Average	Poor	Poor
84	Poor	Poor	Poor	Poor	Poor	Poor

1.1

Figure 1.1 shows a spreadsheet of resting heart rates for different age groups and the health of thes subject based on the resting pulse.



1.2 Figure 1.2 shows a graph of target heart rates by age.

the heart closes and circulates blood, and skin becomes more pale as the heart expands. We have constructed an algorithm that can measure these alterations in skin pigmentation wirelessly using standard CMOS sensor technology. From these measurements, we can calculate accurate heart rate averages (bpm) from visual processed data. The human eye is not able to detect the subtle changes in skin pigmentation, however a CMOS sensor is able to do this. In situations that are more complex and include noise, we can utilize MIT's Eulerian Video Magnification techniques to further bring out these changes in skin pigmentation. Then, facial feature detection algorithms can crop out a small portion of skin from the face which can then be run by our algorithm to accurately determine heart rate.

III. Method (Algorithm)

In order to properly calculate pulse from a eulerian magnified video of a person's face several steps need to be taken. The first step is to Figure out where in each frame the face, or magnified skin is so that background noise doesn't affect the results. This doesn't need to be done every frame, but only very so often. The next step is to take measurements of every frame. This creates a somewhat unique fingerprint for every frame. Finally, the last step is to apply statistical analysis to the data from the frames to calculate with a pulse with a degree of certainty.

1. Finding the Face

To begin the algorithm needs to find the face or skin in the video. Usually the subject of a pulse video is not very active, the camera is normally very static. Therefore, this process only needs to be applied every second instead of every frame. To calculate how far into the video it is, the algorithm calculates the current frame number mod the frame rate, or frames per second. If the mod value is 0 then the algorithm searches the frame for skin.

To find the face or finger, or skin cutout in the video, the algorithm uses a common technique of image processing: morphological processing. This is frequently used in image processing to detect components of a binary image. Since this method of feature detection in an image requires a binary image the first steps are just turning the current RGB frame into a binary image.

First, the color image needs to be turned into a gray-scale image which can then be turned into a binary image. To transform a color image into a gray-scale image the average red, green and blue values are taken then that value, called the intensity is applied to the red, green and blue values of that pixel making it gray. Once the gray-scale image is computed the binary image can then be obtained through automatic thresholding. The goal of automatic thresholding is to find a good median gray value and split the image based on that value creating an image that isn't too dark or too light. To begin , the first threshold value is set at 128 and the image is computed. The mean tint of the darker gray pixels is then calculated as well as the mean tint of the lighter gray pixels. These two values are then averaged to calculate the new threshold. This process is repeated until the new threshold and old threshold only differ by 1 or less. Now that the best fit threshold has been discovered the algorithm turns the image from the frame into a binary image. Any gray pixel lighter than the threshold value goes to white and any pixel with a darker value turns to black. Finally, morphological processing can begin.

Although the binary image created from automatic thresholding is a good representation of the skin we need to measure it may leave out some key skin pixels that were on the border of the threshold. To include those pixels in the pulse calculations and get a more accurate reading morphological erosion needs to be achieved. Morphological erosion begins by creating a second matrix, B, in addition to the binary image that is 3x3 and every value is 1 in the matrix. After, every white pixel in the binary image matrix, A, is iterated over. At each pixel a binary OR operation is performed with the mask matrix, B. The result is that the white area expands to fill in the gaps left from automatic thresholding. The final binary image is saved as a bitmap file such as those in Figure 2.1 and 2.2 for evaluation later. Meanwhile, the eroded binary matrix is saved and updated every second so that the fingerprinting step can properly be computed.

2. Fingerprinting

Before jumping into the fingerprinting step of the algorithm the mystery what a frame fingerprint really is needs to be addressed. If every framed were masked by the binary image from the last step then the colors shown in each frame will be fairly similar but all slightly unique due to the constant change in skin pigment when the heart beats as well as movement of any lighting source in the video. The colors of every pixel in every frame are normally expressed in red, green and blue values but these can be converted into hue, saturation and intensity values. It is in these values that the algorithm can properly evaluate each frame.

Every time a frame occurs the color values are

2.2

2.1

Figure (2.1) represents the morphological erosion result on a subject's face, while Figure (2.2) represents the same morphological erosion on a video cropping of a different subject's forehead. The top black parts of the mask is where the hair casts a shadow making the skin difficult to read.

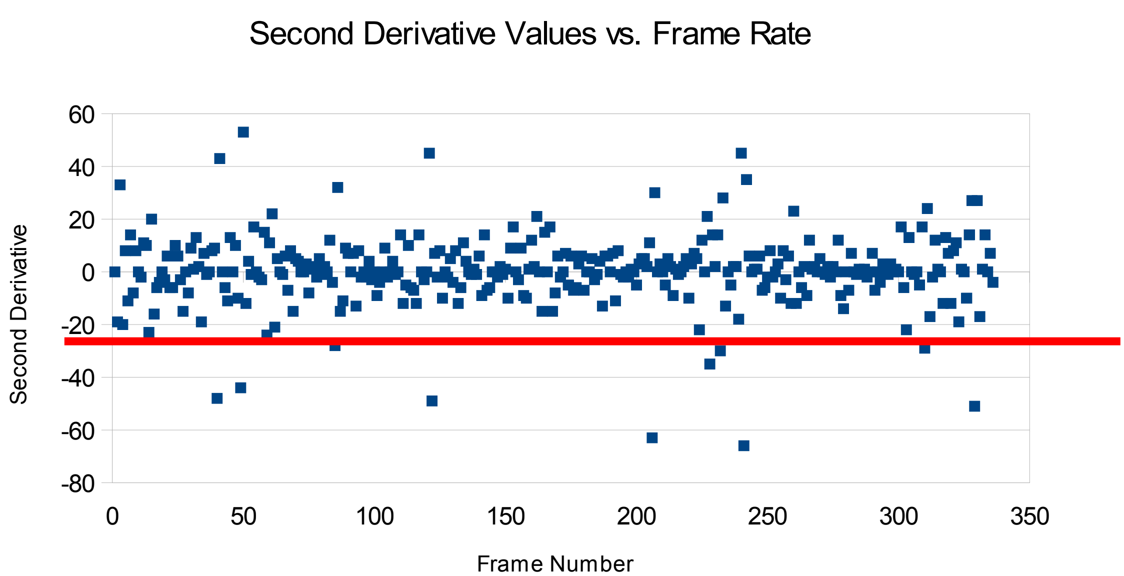
converted into the HSI values and averaged together. The result is three unique properties that are shown in the mask of that frame. Each mean value is pushed into its corresponding vectors. Once the image is done running one vector holds all the averages of hue in every frame, another the saturation averages and a third the intensity averages. These vectors can then undergo statistical analysis to finally determine the subject's pulse.

3. Statistical Analysis

The last and most important and complicated step of this algorithm is statistical analysis. Of the three vectors of data obtained from the last phase of the algorithm at least one of them will contain a sin wave with a frequency equivalent to the pulse of the subject. The objective of the analysis is to find which one and what the frequency is despite lighting noise and movement. Since the changes in hue, saturation and intensity are all pretty faint it is important not to look at the values but the changes in these values, or the derivative of each vector.

The first step of each vector analysis is to calculate the derivative, or change in value over change in index ($\Delta Y/\Delta X$). Although this first derivative may be somewhat useful in calculating pulse second derivative is where the real gold is. At the top of each of these sine wave peaks lies a heart beat. These peaks are when the skin is most red, and actually represents a heart beat. To find these peaks the second derivative is more helpful than the first because what the algorithm really needs is concavity; even more specifically, it needs to find concave down curves. Since extremely low second derivative values are the most extreme, well-defined "heart beat frames" the algorithm computes the second derivative. Once this is finished the algorithm needs to Figure out what constitutes a very low concavity value.

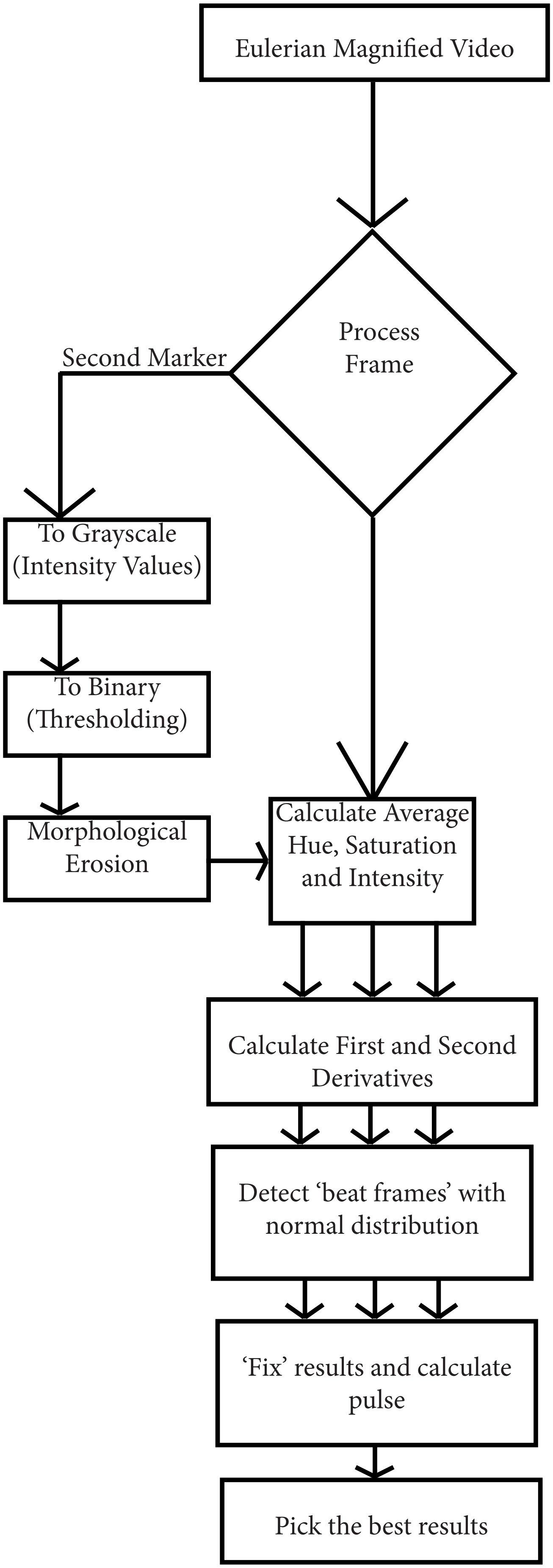
Although this is easy to see by the human eye when looking at a chart the computer cannot "see" these low values. This is where the statistics come into play. The second derivative vector contains mostly of 0s,1s, and



3.1

Figure (3.1) as shown above shows all the second derivative values as the video progresses. The red line is used to represent the threshold of 2 standard deviations below the mean.

Figure 4.1



-1s due to the very small changes in skin pigment, even after eulerian video magnification. However, as seen in Figure 3.1 there are clearly outliers. If treated like a normal distribution of random numbers then a bell curve could be applied and outliers can be picked out. Therefore, the next step is to calculate the mean and standard deviation of the second derivative values. Now we can estimate at which frame numbers heart beats occurred. Every frame number that has a second derivative value lower than 2 standard derivatives below the mean can be considered a concave down outlier and a “heart beat frame.” The numbers of these “heart beat frames” are then put into a new vector to have some final adjustments done to get more accurate results.

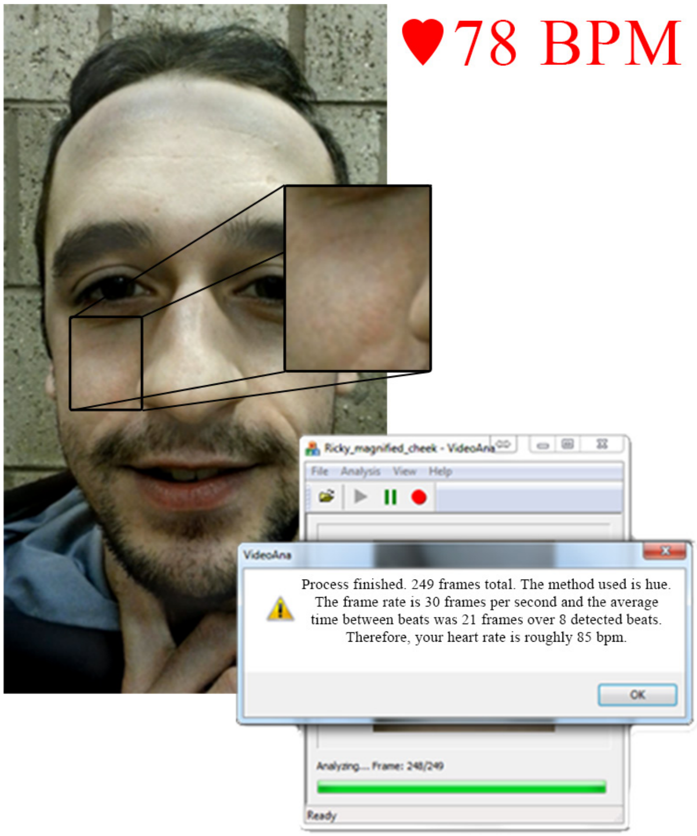
Now the algorithm is coming into its final stretch. Two metrics are key in shaping the vector of “beat frames.” A human heart has a healthy resting rate of anywhere between 50-70 bpm and an active pulse of up to 200 bpm. Based on frame rate the algorithm must compute the minimum number of frames needed for two beats to be only as close together as 200 bpm as well as the maximum number of frames needed for two beats to be at most 50 bpm frames apart. In a standard 30 fps video this means that beats cannot be any farther than 36 frames apart and no closer than 9 frames apart. The vector of “beat frames” is then iterated through and if any two beats are within that minimum limit then they the lower value is saved and the higher one is removed from the vector as it was probably apart of the same peak wave. Finally, the distances between each beat frame is taken and now stored in one last vector. From here the pulse can be calculated. Ideally the program picked out 10-12 good beat frames that all have about the same distance of frames between them. In these cases the median value of the distances can be taken as the pulse. However, if there are either not a lot of good beat distances calculated or they are very spread apart then the mean of the set of distances is taken. Either way, this final value is then converted from frames to beats per minute using frame rate. The weight, or certainty of this technique is the size of the distances vector, or how many beat frames it detected.

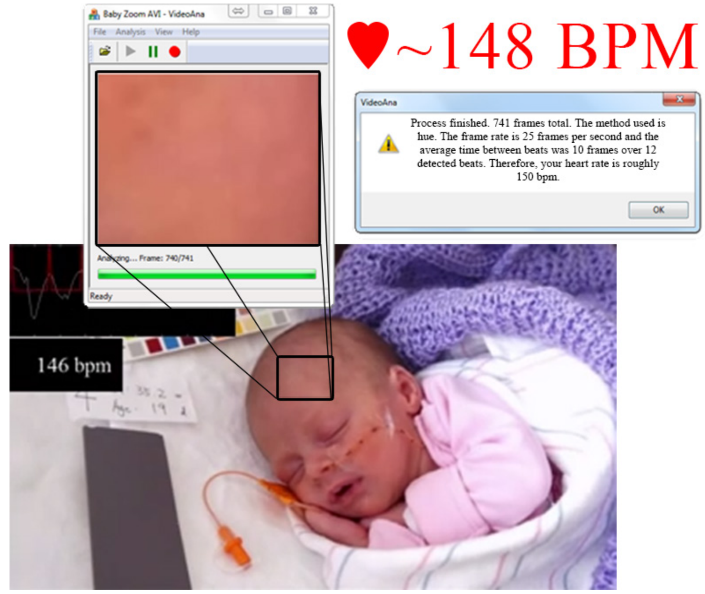
All three data vectors from the fingerprinting stage are run through this analysis and the one with the highest resulting weight or beats detected is the one shown to the user. The entire process can be summed up in Figure 4.1.

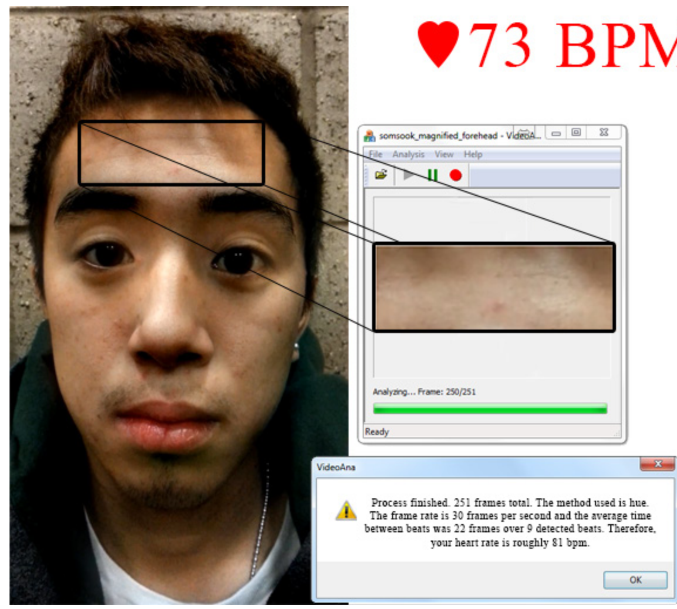
IV. Results

The results of this algorithm showed that it works fairly well when the video is eulerian magnified as well as cropped so that the video is zoomed in on a cheek or forehead. The eulerian magnification necessary to bring out the pulse has an alpha value of 20, low frequency of 5/6Hz, high frequency of 1 Hz, magnification parameter of 6, chrominance Attenuation of 1 and uses the color method. The tiny patches of skin are good indicators for the program as long as shadows and glares don't interfere. The four test subjects shown in Figure 5.1-5.4 have all been eulerian magnified and cropped. In Figure 5.1 the subject's pulse was actually 78 bpm, or about 23 frames between beats at a frame rate of 30 fps. However, the program guessed 85 bpm, or 21 frames between beats. In Figure 5.2 the baby's pulse is around 148 bpm, or roughly 10.13 frames between beats and the algorithm calculated 150 bpm or 10 frames between beats. In Figure 5.3 the subject's frame distance was 24.5 frames between beats while the program computed 22 frames between beats. Finally, the last Figure shows the subject's frame distance should have been 14.3 frames but instead was 13 frames between beats. Overall the four cropped magnified videos had a mean percent error of 7.525% by frame rate. Unfortunately there are some big limitations to this algorithm.

One major restriction is that the program is limited by frame rate. This limitation on how precise the estimation can get is called round off error. Since half frames cannot be calculated the algorithm must go to the nearest frame causing some error in pulse. Another big disadvantage to the algorithm is the necessity for both eulerian magnification method as well as cropping. The algorithm has been tested using videos that have either not been magnified as well as videos that have not been cropped and the program may sometimes get a lucky guess but it has a much higher percent error. Also, light and shadows can cause misreadings because it can hinder the hue, saturation and intensity values. Finally, the program is limited by the camera that takes the video. If the video is blurry background noise can get in the way of the algorithm. If a camera has a slow frame rate then the frame distance error might not be far off but the pulse will fluctuate greatly.







5.1 5.2 5.3 5.4

Figure (5.1) estimated a pulse of 85 bpm, with a distance of 21 frames between 8 beats using hue. Figure (5.2) estimated a pulse of 150 bpm, with a distance of 10 frames between 12 beats using saturation. Figure (5.3) estimated a pulse of 81 bpm, with a distance of 22 frames between 4 beats using hue. Figure (5.4) estimated a pulse of 138 bpm, with a distance of 13 frames between 9 beats using saturation.

V. Future Work

In order for this technology to be commercially available to everyone some improvements need to be made. Starting at the hardware level for life or death emergencies such as in hospitals a camera unit would need to be designed with this algorithm in mind. The camera could still be as small as a CMOS sensor, except it would need to excel at frame rate to ensure more accurate readings. It would need to have facial detection software built in as well as a sharp enough zoom to record video of a cheek or forehead from a far. Internally a strong graphics card combined with a weak processor and a decent amount of RAM is ideal. The graphics card is a necessity for handling the numerous computations that goes with in this case real-time image processing. The processor needs to delegate more than compute. As long as the graphics card does its job well then the CPU won't have to be strained or even that powerful. Finally, the RAM needs to be large enough to handle the large amount of data that comes with video files. If a Wi-Fi chip is also in the unit then a doctor could monitor their patient's vitals from their smart phone while the camera watches the patient closely.

As for the software of the unit there would need to be an automatic flow between eulerian magnification, facial feature detection and the algorithm in this paper, calculating the pulse. There would also need to be adjustments made for changes in lighting and frequent motion tracking. However, once the three algorithms work together then this technique can be simply put into an app for Android or iPhone where you can take your pulse as you run on the treadmill, bike or work out.

VI. Conclusion and Discussion

We have shown that accurate detection and measurement of heart rate average beats per minute (BPM) is possible without the use of direct physical contact sensors through means of visual processing techniques utilizing CMOS sensor technology. With the blend of MIT's Eulerian Video Magnification, a CMOS sensor, facial feature cropping and our algorithm we can detect heart rate in a wide range of circumstance. Our method can be successfully implemented for both long term and short term evaluation and satisfy all of the devices that heart rate information can provide through any other medium. With the ease of access about an individual's personal health that our method will provide to the average consumer, it is our hope that the average person will become more informed and take a more active role in their own health. Our method may lead the way to individual's checking their hearts as often as they would check their cell phone for a message. A possible concern of this technology is that with the elimination of physical contact and the added power of measuring heart rate from a distance, can we maintain our heart health privacy? This may open up the door to being secretly scanned by unauthorized parties who wish to obtain our own personal health information for unethical reasons. What if an individual was participating in a job interview and the interviewer was secretly recording the interviewee's heart rate for malicious intentions? The issue arises of how can one keep this information private with such technology easily able to detect one's heart rate.

VII. References

[1] G. Abowd, A. Bobick, I. Essa, E. Mynatt, and W. Rogers. The Aware Home: Developing Technologies for Successful Aging. In AAAI Workshop and Automation as Care giver, 2002.

[2] Principles of pulseoximetry Clinicalmonograph,1991.R.

[3] R. N. Anderson. Method for Constructing Complete Annual U.S. Life Tables. National Center for

Health Statistics. Vital and Health Stat,, 2(129), 1999.

[4] Wu, Hao-Yu, Michael Rubinstein, Eugene Shih, John Guttag, Frédo Durand, and William T. Freeman. “Eulerian Video Magnification for Revealing Subtle Changes in the World.” *ACM Transactions on Graphics* 31.4 (2012): n. pag. Print.

[5] Yin, Lijun. “Color Theory.” Color Theory. Binghamton University, Binghamton. 8 Oct. 2013. Lecture.

[6] Yin, Lijun. “Morphological Image Processing.” Morphological Image Processing. Binghamton University, Binghamton. 29 Oct. 2013. Lecture.

[7] Zhu, Mingliang. “An Easy Video-processing Framework by Grabbing Frames as Bitmaps Using DirectShow.” *CodeProject*. CodeProject, 5 Mar. 2008. Web. 15 Dec. 2013. <http://www.codeproject.com/Articles/14689/An-Easy-Video-processing-Framework-by-Grabbing-Fra>.