

Design Document

Core Body Temperature Estimation to Detect Ebola Virus Disease

CS 461, Fall 2017, Group 34

Claude Maimon, Brian Lee Huang, and Bianca Beauchamp



Abstract

The end goal of this project is a research paper as well as a prototype of the software. The paper should outline the problem that the project is trying to solve, the steps taken to solve the problem and how successful the implemented solution was. The prototype of the software as well as the information in the paper should allow the project to be continued by someone else. The whole process should be explained in detail allowing whoever wants to continue the project to continue without any problem. The main body of the research paper will be about the program that is developed to predict a person's core body temperature using a thermal image. The program will first be able to extract data from a thermal image. The data from the image will come from the top half, and then be narrowed down to the head of the person. It will then use the data to create a mathematical model that will predict core body temperature. A high accuracy rate is not strictly required as this is an exploratory project. The main goal is to determine if this method will be effective to detect a elevated core body temperature. If a high accuracy rate is achieved then there this method should continue to be perfected. If the outcome is a low accuracy rate then this method may not be best and a new method should be considered.

CONTENTS

1	Introduction	3
2	Importing images and data	3
3	Images into right format	3
4	Pixel selection	3
5	learning 1	4
6	learning 2	4
7	learning 3	4
8	Statistical Analysis	4
9	Production Model	4
10	User Interface	5
11	Conclusion	6
	References	6

1 INTRODUCTION

There are four major parts to this project, image processing, machine learning, production and evaluation. The image processing portion will import the thermal images from the camera, format the images, select necessary pixels, and summarize those pixels into a single temperature value. The machine learning portion of the project will create a mathematical model to represent the relationship between skin temperature and core body temperature and then test this model on different sets of data. The production portion will use the model that has been created to produce an estimated core body temperature from the estimated skin temperature and then produce a true or false output as to whether or not that subject has a fever. The evaluation portion will evaluate how accurate the true or false output of the production portion is. The machine learning, production and evaluation portions will have user interfaces. The first piece of the project that will be developed is the image processing portion since both the learning, and production code rely on processing the thermal image to get a skin temperature. The second piece that will be worked on is the machine learning portion since the model needs to be created and tested before it goes into production. The third piece that will be done is the production portion and the fourth piece that will be done is the evaluation portion.

2 IMPORTING IMAGES AND IMAGE DATA

The first step for image processing is importing the image from the camera. It's important that the image will stay in the right format to keep the pixel data. This means that the image needs to be in a format which maintains the temperature data. The FLIR Tools software will be used to achieve this.

FLIR Tools is a software for importing and analyzing images from FLIR cameras. The software can be used to import images to a personal computer, search the image library using various filters, store search criteria and manipulate images. The software is free to use and has a simple installation process. FLIR Tools is suitable for the FLIR A315 camera. Other than importing images the tool can be used to create PDF reports, add header logos to images and sort the image folder by specific variables.[?]

This software is the best option because it's free and it offers all the features needed for this process. Even though FLIR Tools+ offers more features, it is not necessary and this option will make the project cheaper.

Other than importing images, the core body temperature will need to be collected from many people. These temperatures are needed to compare to the estimated skin temperature from the camera. People's temperature will be collected and stored in a file. Every temperature will be paired with an image, so it is clear which temperature belongs to which person.

The thermal camera may not arrive in time for data collection. If this occurs, black and white images taken from a web camera will be used as mock thermal images. The results from these images won't be accurate but they will make a good placeholder so that work can be done.

3 FORMATTING IMAGES

In order to process the thermal images, they have to be in the right format. After importing the images from the camera to the computer, the program will transfer the images into a two-dimensional array format. This format will allow for easy image processing. Every element in the two-dimensional array will hold a value of a pixel in the image. This structure will allow for easy processing of the pixel values.

The program will use OpenCV with Python to manipulate the image. OpenCV (Open Source Computer Vision Library) is a free open source computer vision software. It has interfaces for C++, C, Python, and Java and it supports various operating systems. OpenCV's library has more than 2500 algorithms which offer many features. Some of those features are facial recognition, gesture recognition, motion understanding, bio-medical analysis and more. The software is used all around the world, has a strong user community and offers technical support. OpenCV is a good option for manipulating the image because it is free and has a large user community. It has shown to have the best performance in comparison to other image processing libraries and there are many easy to understand tutorials available.[?]

4 PIXEL SELECTION

In order to get a good temperature estimation from the image, only the pixels from the warm parts of the subject's head are needed. The program will go through the pixel values in a two-dimensional array and select specific pixels. It will only select values that are higher than 98.8 and lower than 105. The pixel values might not be represented as degrees. If they are not represented as degrees the pixel values that represent 98.8 and 105 degrees will be found and those values will be used instead. This approach will work because by only selecting these values it will, in theory, isolate only the warm parts of the subjects's head and it will leave out the cooler parts of the head as well as the background behind the subject.

Since this is a primarily a research project this method will be tried first, but if it fails a different approach to selecting the pixels may be taken. Another approach may be to isolate the subject's head from the picture before getting the pixel values. This way the background would be removed prior to selecting the warmest parts of the face which would in theory eliminate the chance of the background being included in the data.

5 SUMMARY STATISTICS

Once the pixels of skin have been isolated from the image, they need to be simplified into a single value that is a good representation of the entire set of data. This single value would represent the temperature of the subjects skin. This is important because the value will be used in the model to relate to core body temperature.

The mean will be used for the summary statistics because it is the most accurate representation of all the data points and it is simple to find. In this case the mean used would be the arithmetic mean which is the average of the entire data set. The arithmetic mean is found by adding up all of the values in a data set and then dividing by the number of data points in the data set. The mean is the best used as a representation of the entire data set because it minimizes the sum of squared deviations from the typical value, meaning that it is a value that is the closest to all values in the set. This will be important because the value from the summary statistics will be what is used to relate the data from the camera to the measured value in the model.

6 MODEL

The condensed data needs to be related to the measured value in order to produce an equation that represents the relationship between skin temperature and core body temperature. The value produced by the summary statistics represents the skin temperature of a subject and the measured core body temperature of the same subject will be provided. These two pieces of data will be taken from many subjects and then used to create a model. The model will

then be able to take just the summary statistics, which is the subjects estimated skin temperature, and produce the core body temperature.

The model must be able to represent the relationship as accurately as possible. This means that the type of model used will need to be determined by trial and error since there is no known relationship. The order in which the different models are tried should go from least complex to most complex.

The simplest model is a linear regression model. This model is used for modeling the relationship between one dependant variable and one or more explanatory variables. This type of model is typically used for predictions, forecasting and error reduction. The predictive model is made by finding the equation of the line of best fit for a set of dependant and explanatory variables. Then the model can be given an explanatory variable and predict the dependant variable.

A slightly more complex model is a polynomial regression model. This model is used for modeling the relationship between dependant and explanatory variables as an n th degree polynomial. The model is made by fitting a nonlinear relationship between explanatory variables and the corresponding conditional mean of dependant variables. Polynomial regression models are used for nonlinear phenomena.

A even more complex model is a ridge regression model. This is a type of regression model that is used when there are explanatory variables that have a high correlation. It is very similar to the linear regression but it is more complex since in order to account for the explanatory variables with a high correlation it uses the prediction errors.

The linear regression will bet the best model to try and test first because it is the simplest of all the models. If the linear regression does not produce a high enough accuracy during testing, the next model tried and tested will be the polynomial regression since it is slightly more complex. If the polynomial regression is not accurate enough, then the ridge regression will be tried and tested. If all three do not produce the accuracy that is desired then other models besides these three may need to be considered or it may not be possible to produce an accurate relationship between skin and core body temperature.

7 MODEL TESTING

Once the model has trained on the set of data that has both the subjects estimated skin temperature and their measured core temperature, it needs to be tested on a new set of data to determine how accurate the model is. This will be done by providing the model with only the estimated skin temperature and comparing the core temperature the model calculates to the measured core temperature. To quantify this comparison it is best to calculate the error between the calculated core temperature and the measured core temperature. This will need to be done for a large set of data to get a good idea of how well the model is working. The error calculation will provide valuable information that will allow for an accurate evaluation of any error in the relationship between the calculated core temperature and the measured core temperature.

Absolute error will be used for finding the error of this model. Any error at all is detrimental in this case and absolute error provides a simple evaluation of if the model is accurate or not. Absolute error averages the size of the error and it weights each error the same. This is done by subtracting a measured quantity from a calculated quantity and taking the absolute value of the result, repeating this for all sets of measured and calculated values, adding all of these values together and dividing by the number of sets.

8 EVALUATION

After the model is created we need to test how accurate the production model is. There are three different statistics that we can use to evaluate how well the product is doing. We can use false positives, false negatives, and accuracy. The rate that we really care about is the percent of false negative. While the other statistics are important, the false negative statistic is critical to keep as low as possible. Sending a unhealthy person home could just result in spreading the disease even further. We would also like to have our accuracy rate as high as possible, as that is the whole point of the project. False positive is the least important statistic as while it is not preferred to have a false positives, it is much less harmful to keep a healthy patient than sending a unhealthy patient home.

To produce data for these statistics we will be using a model to evaluate skin temperature and relate that to body temperature. Core body temperature usually sits around 98.6 °F plus or minus one degree. Anything lower or higher is not normal, and results in either hypothermia or a fever. For this project we only care about the higher range, so any temperature under the range can be considered safe. Anything higher can be considered in the fever range, so these temperatures would not pass. Our model will be based around these thresholds, so anything in the range of core body temperature and below would be considered safe. Anything higher would be considered unsafe and would be considered symptomatic with Ebola. To gather data for these statistics we simply create a running tally of the predicted core body temperature, the real core body temperature and if the model was correct or not. Then this table would simply be outputted as some text file that we could easily analyze and look at the rates.

To minimize our false positive rate we can use a receiver operating characteristic curve to evaluate how well we are doing, and to adjust the threshold as necessary. This curve is dependant on the false positive rate, and the true positive rate, and changing the threshold will allow us to modify those specific rates. Using the curve and changing the threshold will allow us to modify our program to be as accurate as possible. The ROC curve however does not take false negative into account, so we would need to find another method along with the ROC curve to make sure the false negative rate is kept at its absolute minimum.

9 PRODUCTION MODE

The production mode is the third piece of the overall project. The production mode will use the model produced by the learning portion of the project to analyze the image. The output of the production mode will be the estimated core body temperature produced by the model as well as a true (fever) or false (no fever) statement. This portion of code will be written in python like the other pieces of the project to eliminate any need for cross communication of languages. Although this is an important piece of the project most of it can only be done after the image analysis portion is complete. The production code also requires the model created by learning portion of the project. The production mode will use the image processing portion of the project to get the data from the image and then it will use the model created by the learning portion of the project to estimate the core body temperature of the subject.

The production mode should be accurate as possible since it will be used to predict a patient's core body temperature. This would mean keeping the false negative rate as low as possible. However, since this is primarily a research project it is not realistic to have a perfect program in such a small time frame. So the goal that has been established is a 40

10 USER INTERFACE

Once the project gets started a user interface is going to be one of the first things constructed. The initial user interface will have a command line user input and print out the results to the screen. The user interface will be created using

python's file input and print statements. It will be up to the user to properly parse the input files and properly input the relevant information. This simple interface will be used for most of the project, if not the whole project. This will depend on how accurate the model becomes and how much time is left. If the model is very accurate and there is time left a nicer user interface will be created. As the project becomes more and more complete the user interface should also improve with the project. The user interface will slowly develop as the project continues to progress since more inputs and outputs may become necessary.

If there is time to improve the user interface a GUI (graphical user interface) will be created. Instead of using command line for input and output the user interface will allow the user to select the input file using a mouse and keyboard and then allow the user to choose where the output file should go. The output of the machine learning portion should output a text file that contains the model. The output of the production mode should output a temperature that is predicted from the model as well as a true or false statement to the screen.

11 CONCLUSION