Computational Pain Quantification and the Effects of Age, Gender, Culture and Cause

Colin R. Ostberg
Marquette University

COMPUTATIONAL PAIN QUANTIFICATION AND THE EFFECTS OF AGE,
GENDER, CULTURE AND CAUSE

by

Colin R. Ostberg, B.S.

A Thesis submitted to the Faculty of the Graduate School,
Marquette University,
in Partial Fulfillment of the Requirements for
the Degree of Master of Science.

Milwaukee, Wisconsin

May 2014

ABSTRACT

COMPUTATIONAL PAIN QUANTIFICATION AND THE EFFECTS OF AGE,
GENDER, CULTURE AND CAUSE

Colin R. Ostberg, B.S.

Marquette University, 2014

Chronic pain affects more than 100 million Americans and more than 1.5 billion people worldwide. Pain is a multidimensional construct, expressed through a variety of means. Facial expressions are one such type of pain expression. Automatic facial expression recognition, and in particular pain expression recognition, are fields that have been studied extensively. However, nothing has explored the possibility of an automatic pain quantification algorithm, able to output pain levels based upon a facial image.

Developed for a remote monitoring context, a computational pain quantification algorithm has been developed and validated by two distinct sets of data. The second set of data also included associated data for the fields of age, gender, culture and cause of pain. These four fields were investigated for their effect on automatic pain quantification, determining that age and gender have a definite impact and should be involved in the algorithm, while culture and cause require further investigation.

ACKNOWLEDGEMENTS

Colin R. Ostberg, B.S.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ACCRONYMS

AAPM   American Academy of Pain Medicine

AU   Action Unit

FACS   Facial Action Coding Systems

FERET   Face Recognition Technology

HMM   Hidden Markov Model

HSV   Hue Saturation Value

IBCRF   International Breast Cancer Research Foundation

IRB   Institutional Review Board

MPQ   McGill Pain Questionnaire

NRS   Numerical Rating Scale

PCA   Principal Component Analysis

RGB   Red Green Blue

VAS   Visual Analogue Scale

VRS   Verbal Rating Scale

YCrCb   Luminance Color Difference Signal

CHAPTER 1 INTRODUCTION

Pain is a commonly experienced unpleasant sensation that can be caused by both physical and emotional stimuli. It can be something as minor as an inconvenience or irritant, or in severe cases as the Handbook of Pain Assessment says "a serious threat to one's freedom, to the significance of one's life, and ultimately to one's self-esteem." [1]. The American Academy of Pain Medicine (AAPM) gives a variety of statistics related to pain and pain management, including that chronic pain affects more 1.5 billion people worldwide, and at least 100 million Americans [2]. The American statistics indicate that roughly four times as many people suffer from chronic pain than do either diabetes or heart disease, and ten times as many when compared to cancer. However, before looking further at how pain affects people, it is important to define what pain is and how it is typically assessed as well as to define a variety of terms associated with pain.

The Handbook of Pain Assessment defines pain as a multidimensional construct [1] and gives a breakdown of those dimensions and their typical means of assessment. The typical and most comprehensive pain assessment is the McGill Pain Questionnaire (MPQ), which takes means of assessment from each of the four dimensions to give medical personnel a complete picture of the pain the patient is experiencing. The four dimensions are intensity, location, quality and affect. Intensity and location are perhaps the easiest dimensions to define; they express how much pain an individual is in and where the pain is on their body. Three common methods exist to assess pain intensity; the Verbal Rating Scale (VRS), the Visual Analogue Scale (VAS), and the Numerical Rating

Scale (NRS). All three of these methods ask the patient to report their own pain ranging

on a scale from no pain to the worst pain imaginable. VRSs are given verbally and

present the patient with a list of adjectives which increase in perceived pain intensity.

NRSs are mostly interchangeable with VRSs, as they simply ask a patient to supply a

number on the same relative scale. VASs present a patient with a series of images

corresponding to more and more pain, and ask them to select the image that corresponds

to their pain appropriately. Again, this method can be mapped to an NRS with the same

scale. A sample VAS is provided in Figure 1.1.

Figure 1.1: A Sample VAS [1]

Pain location can be assessed via two methods, a simple verbal question can be

presented to the patient, asking where the pain is located. Otherwise, a pain drawing can

be given to the patient, who is then asked to give the location(s) of the pain based on the

drawing. A sample pain drawing is provided in Figure 1.2.

Figure 1.2: A Sample Pain Drawing [1]

The remaining two dimensions of pain can be more difficult to define. Pain quality refers to how the pain feels, independently of how much it hurts. These can also be called sensory qualities, and are assessed with several descriptors, such as sharp, hot, dull, cold, sensitive, itchy, deep and surface. It is not unknown to assign an NRS to each of these descriptors, having patients rate each descriptor on a numeric scale, but this is not a common practice. The final dimension, pain affect, is the most intangible dimension but also the most impactful. Pain affect measures how the pain someone is experiencing affects them on an emotional level. It is important that distinguish that this is not emotional pain, but rather how the patient is coping with the pain, how disruptive the pain is to their everyday life, how bearable the pain is and how the pain is affecting their emotional wellbeing. Pain affect is correlated with pain intensity, although different people are able to deal with it differently. For example, someone dealing with consistent, high levels of pain will be more likely to be affected more than someone with low, intermittent levels of pain, but this is not necessarily the case. Pain affect is typically

assessed via a VRS, although VASs also exist for assessment. A sample 15-point VRS presents the following descriptors to the patient: bearable, distracting, unpleasant, uncomfortable, distressing, oppressive, miserable, awful, frightful, dreadful, horrible, agonizing, unbearable, intolerable, and excruciating.

Two more important definitions associated with pain are acute pain and chronic pain [2]. Acute pain is the typical pain people feel when injured, it is relatively brief and typically milder, although it can reach the heights of pain intensity scales. Chronic pain, however, is pain that lasts for weeks, months or years. The causes of chronic pain are incredibly varied, everything from a sprained back to cancer. This work will focus predominantly on chronic pain. It is also important to note the subjective nature of pain [1], each individual's pain intensity and pain affect can differ dramatically, regardless of the similarities of the causes.

While individuals experience pain individually, it is still something that affects a significant portion of the population, is unpleasant, and costs society as a whole. As mentioned previously, roughly 100 million Americans suffer from chronic pain [2], meaning that a third of Americans experience joint pain, swelling or a limitation of movement at any given time [3]. Pain also puts a substantial burden on our society due to the inability to perform daily activities, loss of work productivity [4] and increased healthcare costs [5]. The AAPM gives these costs as $297-$336 billion per year and $261-$300 billion per year respectively, giving total societal costs of pain as $560-$635 billion per year in 2010 dollars [2]. In order to combat this, various health agencies and regulatory bodies have mandated that pain be assessed upon contact with a healthcare provider [6, 7, 8, 9]. In regards to quality of life, 51% of patients reported that they have

little to no control over their pain, and 59% reported an impact on their quality of life [2].

Based on the numbers provided by the AAPM, that would indicate that between 50 and

60 million Americans have their quality of life affected by chronic pain, and between 750

million and 900 million are affected worldwide. That said, continuous monitoring of pain

intensity in intensive care units improves patient outcomes and quality of life

tremendously [10]. This point and the following quote are what this work strives to do:

> "Pain is a dynamic, developmental process, not a single event or simple quantifiable product…Thus, when we and others talk of 'objective' measures or 'quantifiable' indices, the reader should understand that we do not intend to depict pain as a static, all-or-none, uni-dimensional, body-centered occurrence that exists somehow independently of time, place, the patient's states of consciousness, or the observer's presuppositions. We have elsewhere noted: 'As pain assessors, we are coparticipants, not merely observers and, therefore, although there is no single best way to interpret pain, we can probably serve our patients better if we acknowledge that we are jointly engaged in creating the pain dimensions we seek to measure.'" [1]

The purpose of this work is the development of a computational pain intensity

algorithm based on facial images. It focuses primarily on chronic pain, and strives to

provide additional, easy to obtain information on pain to medical personnel, allowing

them to provide better care, pain management, and hopefully improve the patient's

quality of life. This algorithm is able to take only a facial pain image as input and

determine the pain level of a patient accordingly. Furthermore, this work will investigate

whether the age, gender, pain location or culture of a patient has an effect on

computational pain quantification.

The majority of this work has been done in a Mobile Health, or mHealth, context

for several reasons. First of all, the mHealth field is growing at a tremendous rate, and it

is projected that mHealth monitoring could save nearly $200 billion in the next 25 years [11]. Secondly, while this computational approach may not be immediately applicable within a medical facility, it would be in a remote monitoring scenario. Patients dealing with chronic pain who are not hospitalized could communicate easily with their health care providers and provide them with daily or even more frequent data points. While pain management is important for all chronic pain patients, some ailments place particular stress on pain management in standard treatment plans, such as sickle cell disease [12]. In these scenarios, providing medical personnel with daily updates on pain levels has the potential to increase the patient's quality of life, and possibly help with treatment while maintaining a patient's independence. All of that said, it is this author's dream that health care facilities may someday employ a pain monitoring system evocative of modern day vital sign monitors as standard equipment in patient rooms.

The nature of this work naturally gives rise to several criticisms that demand answers. First is the fear that this work aims to replace medical personnel in the pain management process. Medical personnel are instrumental in pain management. As quoted above, "As pain assessors, we are coparticipants, not merely observers and, therefore, although there is no single best way to interpret pain, we can probably serve our patients better if we acknowledge that we are jointly engaged in creating the pain dimensions we seek to measure." [1]. This work does not in any way mean to replace or remove medical personnel from pain management. Instead, the goal is to provide medical personnel with additional information to allow them to make more informed decisions. In remote monitoring contexts, providing daily information to medical personnel gives them more information than they would typically receive at a scheduled appointment with a patient,

so that in those scheduled appointments medical personnel can make better decisions with more information. Furthermore, it has been shown that in Bangladesh, a developing country where healthcare is not as readily available to the general populace, breast cancer patients tended to overstate their pain intensity levels, universally reporting a pain level of 10 on a 10 point scale, but when given a mobile device and told to submit pain values daily, their pain levels evened out to more truthful levels [13]. It was believe this was because in part the patients wanted to ensure they got appropriate medical attention from their doctors, but when they knew they were receiving medical attention on a daily basis this worry decreased, allowing them to give more accurate data to doctors, which in turn allowed them to make more informed decisions. This can be generalized to other developing countries, particularly as according to a recent UN study, 6 billion out of the world's 7 billion people have access to a cellphone [14], and even further to any remote monitoring patients. Knowing that the medical personnel in charge of your treatment is getting data from you on a daily basis, while still allowing you to live your life as independently as possible, should help to assuage fears about proper treatment and in doing so, help ensure that data is accurate. This however, does bring up a second major criticism; the question of why facial images are necessary. The study conducted in Bangladesh mentioned above assessed pain via a smartphone application that recorded pain as a digital NRS. The question that needs to be answered is what can a facial pain image accomplish or accomplish better that an NRS does not? Both accomplish the same goal of providing more pain information to doctors. The difference is in how that goal is accomplished. One goal of paramount importance to this work is the improvement of patient quality of life, which means improving the amount and quality of data available to

doctors at minimum effort to the patients. The study in Bangladesh found two key points in this regard. First, most participants did not fill out the NRS themselves, instead they had family members or friends do it for them while they relayed the answers verbally. As this work requires only an image taken, this in essence removes that step, as only a daily picture has to be taken. Secondly, when patients did fill out of the NRS themselves, many of them encountered difficulties in inputting proper data due to tremors, caused either by age or pain. When patients submitted information themselves, it was found that they did so with an average error of greater than 1.

1.1 Contributions of Thesis

The contributions of this thesis are threefold. First and foremost, this thesis will provide a novel approach to computational pain quantification. Secondly, this thesis will investigate the effects of age, gender, culture and pain location on this computational pain quantification method. Finally, this thesis will provide two distinct datasets of facial pain images with correlated pain intensity values, both of approximately 400 images, databases that do not exist elsewhere.

1.2 Organization of Thesis

The rest of this thesis will be organized as follows: Chapter two will investigate existing work in the fields of computational facial, emotion and pain recognition. Chapter three will provide an in depth look at how the computational pain quantification model was developed. Chapter four will investigation the additional fields of age, gender, culture and pain location and their effects on pain intensity detection. Chapter 5 will conclude this thesis.

CHAPTER 2 RELATED WORKS

Before delving into computational methods, the choice of using purely facial

images for the work of this thesis, as well as why it was selected for existing

computational emotional and pain detection methods, must be investigated. From a

medical context, the face is recognized as a subjective conveyor of emotion and is a tool

that can be used to gauge the intensity of a subjective experience [15]. It has been stated

that only 7% of a face-to-face message is conveyed with linguistic language, while 38%

is due to paralanguage and 55% of message is conveyed via facial expressions [16]. The

Handbook of Pain Assessment extends this to pain specifically, indicating that the face is

an ideal conveyor of not only emotion, but pain as well [17]. In fact, the three

psychologists who authored that section of the Handbook of Pain Assessment, Dr.'s

Craig, Prkachin and Grunau, will be cited several more times throughout this chapter

from references dealing with computational emotion and pain recognition, indicating

their support and collaboration in the development of such methods.

**2.1 Existing Facial Detection, Facial Recognition and Emotional Recognition
Methods**

A significant amount of research has been done in the fields of computational

facial detection, facial recognition and emotional recognition methods. In this context,

computational facial detection refers to any method or algorithm performed on a

computer that is able to identify a face using an image as input. This is also commonly

referred to as face segmentation. Computational facial recognition builds upon facial

detection, typically utilizing facial detection methods as a preprocessing step to the

image. As such, most facial recognition algorithms make the assumption that the image to be analyzed includes a face of some sort. Another typical assumption is that the face to be analyzed is unobstructed and forward facing. While work has been done on facial images where the face is in profile or at other angles has been done, this is not the focus of this thesis and is not a necessary precedent set in the field. These facial recognition algorithms are designed to recognize an individual based on the input facial images and typically involve a comparison between the input image and other images that the software has access to. Finally, computational emotion recognition developed tangentially to facial recognition and utilizes some similar techniques, and also has the tendency to make similar assumptions about the input facial images. Emotion recognition is also commonly called emotion detection or expression recognition and expression detection. Emotion recognition is able to use the input facial image and identify an emotion. There are two general methods of doing so. The first method selects an emotion from a list of available emotions. They second method focuses only on one emotion and detects the presence or absence of said emotion. The second method is commonly referred to by whatever emotion is being detected, i.e. a method that detects the presence or absence of pain would be called a pain recognition method. This section will investigate facial detection, facial recognition and emotion detection methods with the exclusion of pain recognition methods, those methods will be investigated in depth in another section.

2.1.1 Skin Color Facial Detection Method

The skin color method is a commonly used and widely accepted facial detection method. Traditionally this method uses three different color spaces, Red Green Blue

(RGB), Hue Saturation Value (HSV), and Luminance Color Difference Signals (YCrCb)
[18]. The Red Green Blue color space is the typical color space associated with coloring
and pixels, each pixel has three values associated with it which correspond to the amount
of the colors red, green and blue present in the pixel. It is believed in the field that the
RGB color space is unsuitable for use in the skin color method due to the fact that human
skin has a wide range of colors, it is unreliable in different lighting conditions [19], and it
is a difficult color space to execute some image processing algorithms in [20]. HSV
places less emphasis on the actual color, corresponding to two components, hue and
saturation. Hue is the pure color, while saturation is the purity of that color present, or the
amount of white light mixed with the hue. Value is indicative of brightness. This color
space is used commonly for image analysis [18]. YCrCb is similar to HSV, as it too has a
luminance or brightness component, Y, but it handles colors differently. Cr and Cb are
the color difference signals, which are derived from the RGB values. This color space is
used extensively in digital video [20].

Both HSV and YCrCb have been shown to be reliable in terms of the skin color
method, so the first step to this method will be to convert traditional RGB colors to one of
these two color spaces, or a combination thereof. Once in these color spaces, identifying
skin is relatively easy, as human skin, regardless of ethnicity, occupies a relatively unique
range of values within each of these color spaces [20]. Once skin pixels have been
identified, it is simply a matter of segmenting the region containing skin off from the rest
of the picture. As one could reason out, this does mean that this method makes the
assumption that a face is the only area of skin within a picture, so it has its limitations.

However, due to the typical assumptions that a facial image will include a forward facing face and includes one face, this method is widely accepted as useable in such a context.

2.1.2 Eigenface Facial Recognition Method

The Eigenface Method is by far the most widely used method for facial recognition. Its influence in the field does not stop with facial recognition either, but spreads to emotion and pain detection. While several additions and modifications exist to the method, they all revolve around Principal Component Analysis (PCA). PCA is a generic algorithm that does not necessarily need to apply to faces or any sort of facial recognition, it is a statistical and data compression technique with applications in image compression, as well as face and emotion recognition [21].

The technique consists of six steps. The first step is data collection, some data must be present to perform PCA on. Once data has been collected, the last part of this step is to calculate the dimensional means, meaning that for every dimension that the data has, a mean must be calculated for that dimension. The next step is to subtract the dimensional means from each dimension at each data point. This yields what is called the data adjust. Step three calculates the covariance matrix of the data adjust. The covariance matrix is simply a matrix made up of every combination of covariance's, or variances between two dimensions, for the entire dataset. So if one has n dimensions, the covariance matrix will be of size nxn, representing every possible pair of dimensions. Step four calculates the eigenvectors and eigenvalues of the covariance matrix. These are principal components of the data, meaning that the strongest patterns and relationships between the data are represented by the eigenvectors. In fact, the most

important and strongest component is the eigenvector with the largest eigenvalue. Similarly, the second largest eigenvalue's corresponding eigenvector is the second most important component, and so forth. The next step then, involves selecting which components to use. All of the components can be used, but the lower components can be eliminated with little loss of data, as they are the least important components. Whichever eigenvectors are selected, whether it be all or a subset, are referred to as a feature vector. With this feature vector, the new data set can be derived in the sixth and final step. The new data is equal to the transpose of the feature vector multiplied by the data adjust. This puts the data in terms of the principal components, allowing for easier analysis [21].

When applied to facial recognition, however, PCA becomes a machine learning tool rather than an analysis tool. It is used in the 5 step eigenface algorithm to train the machine. The eigenface method was originally proposed by Turk and Pentland in 1991 [22] .The first step is to create what is known as the facespace, this is done simply by performing PCA on a set of images denoted as the training set. The resulting eigenvectors, called by the titular name of eigenfaces, make up this facespace. Sample eigenfaces, created with real data from the Longitudinal Study, which will be detailed in chapter 3, are given in Figure 2.1.It is generally assumed that the training images are all of isolated faces. These training images are then projected onto the facespace. The second and third steps are to prepare new images for comparison with these projected images. Step two subtracts the mean image, calculated during PCA, from the new input image, while step three projects that difference onto the facespace. Step four compares the distances between the training projections and the new projection, and step five

finds the minimum distance. It is then assumed that the new image is most closely associated with the image that it is closest to, so from the machine's perspective it takes on the characteristics of whatever training image the new image is closest to. Thus, for facial recognition, the new image is recognized as belonging to the individual who is represented on the training image closest to the new image [23].

Figure 2.1: 6 Eigenfaces

The eigenface method is used widely in the facial recognition literature with marked success for several reasons. For one, it is incredibly accurate, the first iteration of the eigenface method reported a 96% success rate [22]. Additionally it is incredibly modular, and can be used in conjunction with several other methods for increased accuracy, such as neural networks [24, 25, 26] and eigenfeatures, which perform the same eigenface algorithm on smaller features of the face such as on the nose or mouth

[27] in the case of facial recognition. Research to further increase the accuracy of the

basic eigenface method has also been conducted, resulting in a number of preprocessing

techniques, such as using only greyscale images [28], to further increase accuracy.

Furthermore, the eigenface method can be used in near real-time, seeing as the

computationally expensive part, PCA, can be performed while the system has no other

demands placed on it [22].

2.1.3 Evolutionary Pursuit Facial Recognition Method

While technically a different method for facial recognition, evolutionary pursuit is

built heavily on the principles that the eigenface method is built on, namely PCA and

several steps of the eigenface method. Where it differs is that it applies two additional

transformations, the whitening transformation and rotation transformations, to the

facespace, and then applies genetic algorithms to search for solutions. The whitening

transform reduces the dimension of the facespace, while the rotation transformations set

up a basis, which in turn allow the genetic algorithms to search. This method reported as

high as a 92% success rate. There is also no indication given as to if evolutionary pursuit

can be done in near real-time, as the eigenface method can [29].

2.1.4 Facial Action Coding System Emotion Recognition Method

Facial Action Coding Systems (FACS) was one of the first methods used to look

at emotion recognition. It has been researched rather extensively [30, 31, 32, 33, 34, 35]

both as a standalone system and in regards to emotion recognition. It was first published

by Ekman in 1978, where instead of being used for computerized, automatic emotion

recognition, it was designed for readers to teach themselves emotion recognition. It

consists of identifying facial Action Units (AUs), combinations of which make up every

facial expression possible. Ekman has identified over 500 AUs. Figure 2.2 shows a

sample face of fear with detail action units.

Figure 2.2: Sample Fear Face with Action Units [36]

When identifying small amounts of AUs, or a single expression, automatic FACS

performs very well, able to operate in real time [32] and with high accuracy, usually 97%

to 98% [34]. FACS does tend to run into trouble, however, when combinations of

expressions are presented on the same face [34]. FACS locates the features that it needs,

known as feature extraction, and then locally determines the appropriate AUs. The active

contour method is one commonly used method to accomplish this. An active contour,

sometimes referred to as a snake, is a curve with a set of control points, and is used to

find lines, edges or contours. These snakes can isolate local regions, which can then be

analyzed. A common method of analysis is via Hidden Markov Models (HMMs),

although other classifiers can be used [37], such as Markov-chain Monte-Carlo methods

[32], Support Vector Machines (SVM) [33, 38], AdaBoost [33] Neural Networks [34], and multi-layer perceptrons [39].

FACS does have several limitations, namely that faces and the underlying musculature are all different, and therefore the AUs are necessarily slightly different, and due to these different facial geometries, skin colors or brightness, and illumination conditions, in a real world setting much of the feature recognition has to be done manually [40].

2.1.5 Eigenface Emotion Recognition Method

The eigenface method when applied to emotion recognition is very similar in approach to the eigenface method for facial recognition. It uses the same algorithm, but instead of assigning an identity to each image, i.e. a person to identify, an emotion is assigned to each image. When presented a list of 6 emotions to identify, the eigenface method had a success rate between 70 and 80% [41].

Previously, one of the strengths of the eigenface method was its modularity and ability to be readily combined with additional methods to improve accuracy. This is still the case in regards to emotion recognition. Recognition rates of 94-97% were reported with use of Local Gabor Filter Banks and Linear Discriminant Analysis [42], 94% for the Automated Pixel Selection method, which improved by an additional 2% with the use of the facial masks method [43]. This 2% increase with the use of facial masks was confirmed by the same authors but this time using only the basic eigenface method [44]. A generalized eigenspace method based on class features reported anywhere from 45%- 100% recognition [45], and a study also using class features, but applying kernels

reported similar recognition rates [46]. Finally, a method using hierarchical radial basis function networks reported rates between 66 and 75% for lips and 70-100% for eyes, that is, images of only lips and eyes [47].

While not directly applicable, it is worth noting that based on the success of the eigenface method, a performance animation method was developed based on facial expression recognition using eigenfaces [48].

2.2 Existing Pain Detection Methods

Pain detection methods, both automatic and manual, have undergone extensive research amongst a variety of age groups, from neonatal to adult, and also look at patients with cognitive impairments that can not communicate their pain in a traditional manner [49, 50, 51, 52, 53, 54, 55]. The focus of these methods is to identify the presence or absence of pain in the subject, and they do nothing dealing with pain intensity.

2.2.1 Facial Action Coding Systems Pain Detection Method

The FACS methods for pain detection are similar to the generalized FACS emotion recognition. Several AUs have been identified by several sources to be indicative of pain, namely brow lowering, orbit tightening, levator contraction and eye closing [56, 57, 58, 36]. The sources cited give varying success rates, from 88% [58] down to 46% [57].

2.2.2 Eigenface Pain Detection Method

The eigenface pain detection method is similar to both the facial recognition and emotion recognition methods, but instead placing the presence or absence of pain as the

classifiers to be identified with. The eigenface method also retains its modularity when it comes to pain detection, allowing researches to couple it with other methods. Monwar was instrumental in using eigenfaces for pain detection, and most of the cited sources in this section were authored by him and his colleagues. Use of neural networks with the eigenface method gave 86-93% accuracy [59, 60], and use of a face mask gave 91% accuracy [60]. The use of eigeneyes and eigenlips was also investigated by Monwar, where the same approach to eigenfaces was used on eyes and lips. In this case, the eigenfaces alone gave 89% accuracy, eigeneyes gave 82%, eigenlips gave 84%, and a combination of all three gave 92% accuracy [61]. Monwar also extended his work to video based pain recognition [62]. While he did not give an accuracy percentage in this work, seeing as he extracted frames of the video and used those as pictures, it can reasonably be assumed that the accuracy was in line with the rest of his work.

2.2.3 Support Vector Machine Pain Detection Method

In a generalized sense, SVM uses a similar approach to the eigenface method. It takes extracted feature vectors and maps them to hyperplanes, or high dimensional spaces, via a kernel function. The distance between these hyperplanes is then calculated. This is done to train SVM, much like the eigenface method does. When a new image is input, SVM finds the optimal hyperplane, giving the result in a somewhat similar fashion to the eigenface method [63] in that it compares the training hyperplanes to the new hyperplane to get the result. SVM alone gave a 96% accuracy [63], but like the eigenface method, SVM is typically paired with other methods such as Adaboost, Gabor Filters and Hidden Markov Models [58, 36]. Relevance Vector Machine, a Bayesian extension of

SVM, was used to increase classification accuracy over SVM [64]. Relevance Vector

Machine achieved an accuracy rating of 91%.

**CHAPTER 3 DEVELOPMENT OF COMPUTATIONAL PAIN
QUANTIFICATION METHOD**

Of the criteria laid out for the computational pain quantification method first and

foremost is that it must be an entirely automatic method capable of taking facial images

as input and outputting pain intensity values. Both keeping it automatic and limiting it to

purely facial images is for the express purpose of providing as much data to medical

personnel as possible while at the same time minimizing the amount of time and effort

required on both the patients and medical personnel's parts. It must be reasonably

accurate for a remote monitoring context. In initial talks with doctors, it was decided that

an absolute error of approximately one would be appropriate for this context. Because

this is designed to be in a remote monitoring context, runtime is not a necessary attribute,

but reduced runtime is still an influential criterion particularly if future work is to develop

this in or into different contexts. Finally, the method must be robust and able to handle

varying image conditions, such as lighting, image quality and differences in the

appearance of patients. It was also decided at this stage that self-reported pain intensity

values would be taken as accurate, primarily due to the subjective nature of pain. Each

patient will react to different stimulus differently, so the pain intensity values that the

patients reported, whatever they were, were to be taken as absolutely accurate.

3.1 Selection of Existing Methods

The selection of methods to build off of was important. While no computational

pain quantification method has been developed, basing the method off of existing, proven

methods was a logical first step. The eigenface method was selected as this basis for

several reasons. First was for its pervasive success throughout the field. The method has worked reliably and accurately in all areas of facial detection, facial recognition, emotion recognition and pain recognition. Furthermore, the method has proven to be incredibly modular, able to be paired with several other methods for increased accuracy. It was believed that this could possibly be extended to not only increased accuracy, but also to determine pain levels in additional to identifying pain. This was believed due to the fact that images can be paired with classifiers, as in the case with emotion detection, several different emotions could be detected. The natural extension was to apply it to pain levels. Furthermore, this method is fully automatic, has proven capable of processing images and is able to do so at high speeds. Again, while runtime is not a necessary criterion for remote monitoring, it is still beneficial and worthwhile to investigate for possible future applications. Finally, the method has also shown to be able to handle varying image conditions, particularly in differences in the appearance of patients.

3.2 Longitudinal Study Data and Evolution of Basic Method

With a framework for a model selected, the next step was to develop the algorithm. In order to do so, data was required. Fortunately all of the data from the Longitudinal Study was at hand for use in testing and development. With support from the International Breast Cancer Research Foundation (IBCRF) and approval from Marquette University's Institutional Review Board (IRB) as well as Bangladesh's health department, the Longitudinal Study collected data from rural breast cancer patients in Bangladesh [13]. For this study, fourteen terminally ill breast cancer patients were given smartphones equipped with a data collection application. This application collected, amongst other fields, daily pain intensity values and associated facial images.

Unfortunately, due several reasons including patient deaths, only six patients submitted

usable data. From those six patients, 444 facial pain images were collected. Not all

patients submitted the same amount of data, either, some submitted only a handful while

others gave a large amount. Table 3.1 breaks down the distribution of images by patient

and Figure 3.1 details the distribution of pain intensity levels per patient.

Subject	Age	Images
Subject 1	45	14
Subject 2	50	116
Subject 3	42	158
Subject 4	44	12
Subject 5	51	114
Subject 6	38	38

Table 3.1: Distribution of Subject Image Numbers and Ages

Figure 3.1:Pain Intensity Distributions by Subject

This data would be used to develop and test the computational pain quantification method. Initially, the data was divided into two sets, testing and training. 36 images were randomly selected from all patients and pain levels to form the training set, although it was ensured that all pain levels (1-10) were represented, while the remaining 408 images were used as the test set. The initial computational pain quantification method at this stage was very much akin to the standard eigenface method. It relied on PCA to train the method and Euclidean distance was used to determine the distances between training projections and the new input projections. The pain intensity value of the closest image was then selected as the pain intensity value of the input image. This initial method gave a resulting mean absolute error of 2.267, computed from all 408 test images. It is worth

noting that at this stage, a dataset of acted pain images was also tested. This dataset had 36 images in it, and produced a mean absolute error of 2.5. While this dataset was discarded at this phase, it is worth noting that simulated pain was looked into and ultimately rejected.

The next stage of development looked at personalizing the method in an effort to reduce error. The President's Council on Advisors on Science and Technology defines personalized medicine as "Personalized Medicine refers to the tailoring of medical treatment to the individual characteristics of each patient…to classify individuals into subpopulations that differ in their susceptibility to a particular disease or their response to a specific treatment." [65]. Due to the subjectivity of pain previously discussed, it was decided to attempt a personalized approach. Furthermore, the nature of the eigenface method suits it perfectly to personalized medicine, as using training databases comprised only of a single individual reduces noise that would typically be found in training databases comprised of several individuals. While no direct source has made this claim, it can be inferred from statements such as "On the other hand, two images of different subjects should project to points that are as widely separated as possible." [66] among others and the fact that the eigenface method was utilized to differentiate between individual people. As such, with using this method as a basis, it was decided to look into personalized databases. To do this, the Longitudinal Study dataset was divided based on patient, and then further divided into test and training datasets using the same criteria as for the initial tests. Figure 3.2 shows the breakdown of test and training datasets.

Figure 3.2: Breakdown of Individual Training and Test datasets

The patients with smaller datasets obviously had to have smaller training and testing datasets. The results of using individualized datasets gave an average mean absolute error across all six patients of 1.2458, significantly closer to the goal of 1. This was also a marked difference between this method and previous eigenface methods, as previous methods had made no distinction between individualized and group datasets.

The next step was to look at distance measurements. Up until this point, standard Euclidean distance had been used to determine distances between training projections and input projections. Upon looking at several different distance measurements, Angular distance, also known as Cosine distance, was chosen due to its success in high

dimensional spaces [23]. Seeing as every image within the training dataset adds an

additional dimension to the data, due to the fact that there is an eigenface for every

training image, although the two are not related, and each eigenface is perpendicular to

every other eigenface, and each eigenface is itself multidimensional, a significant amount

of dimensions are at play in this method. Angular distance is calculated by $d(A, B) =$

$\frac{A.B}{|A||B|}$. In practice for images, this is computed by determining the frequency in which

particular pixel values appear in an image and comparing those frequencies based on the

presented formula. So d(A,B) becomes the frequency of the pixels that A and B share

divided by the square root of the product of the individual frequencies of the pixels of A

and B. Unlike Euclidean distance however, the shortest distance between two projections

is the maximum cosine distance. That is, the more similar the frequencies are, the more

similar the projections and therefore images are. As such, instead of Euclidean distance

Angular distance was implemented, giving an average mean absolute error of 1.29,

slightly higher than without Angular distance. The results of the Angular distance

measurement, side by side with Euclidean distance, both acted images and real images,

are shown in Figure 3.3.

Figure 3.3: Acted and Personalized with Euclidean Distance and Personalized with Angular Distance

Figure 3.3, in combination with Figure 3.2, are very influential visually, as they heavily influenced the decision to keep Angular distance, despite its overall slightly higher average mean absolute error. Figure 3.2 shows the training dataset sizes. Subjects 1, 4 and 6 all had databases of 6 images, while subjects 2, 3 and 5 had training databases of 36. This was due to the variance in images available for each patient. However, it has been well documented that when using PCA on images, the amount of eigenvectors, that is eigenfaces, and therefore the amount of training images used has a huge effect on accuracy [23]. Figure 3.4 shows the disparity in accuracy between 6 eigenvectors and 36 eigenvectors, determined using the Face Recognition Technology (FERET) database.

Figure 3.4: Accuracy of Identification Based on Eigenvectors **[23]**

As can be readily seen, the difference between 6 and 36 eigenvectors, using 5 and 35 for reference, is huge. A roughly 35% proper identification disparity exists between the two. So because this is a huge factor in proper identification, the decision to keep Angular distance as the distance measurement was based off of the results of subjects 2, 3 and 5. It is worth noting that this disparity wasn't a factor in determining whether or not personalized databases were effective or not, due to the fact that the increase in accuracy when using personalized databases was significant and unanimous. Furthermore, it still applies to subjects 2, 3 and 5, so that decision stands. As such, looking at only subjects 2, 3 and 5, average mean absolute error for Euclidean distance is 1.083 and .872 for Angular distance, which is within the bounds for error put forth at the beginning of development.

3.3 Integration of SVM with Computational Pain Quantification Method

With the mean absolute error under control, it was time to look at distributions. While having a low mean absolute error is important, it was equally important that the output was in the same distribution of the input. The first step was to look at the current distributions without any changes, which are presented in Figure 3.5.

Figure 3.5: Input and Output Distributions for Subjects 2, 3 and 5

Observationally, the distributions appear similar for each subject. To confirm, the inputs and outputs were compared via a Mann Whitney U test. The null hypothesis for this test says that both the samples are from distributions with equal medians and the alternative hypothesis is that they are not. Subjects 2 and 5 failed to reject the null hypothesis with p values of .18 and .13 respectively, while subject 3 also failed to reject the null hypothesis with a larger p value. The results of this test, combined observationally with Figure 3.5, are indicative that the data comes from the same distribution, but could still be improved. As such, it was decided to implement SVM as an add-on to the current method. While this had minimal impact on mean absolute error, it showed a marked difference in distribution, as shown in Figure 3.6.

Figure 3.6: Input and Output Distributions with SVM

Mann Whitney U tests were again performed, resulting in all three subjects passing with higher p values than before. This was more akin to what was expected for the method, low mean absolute error and similar input and output distributions. As such it was decided to keep SVM in the final method.

3.4 Discussion of Computational Pain Quantification Method Results

With the mean absolute error and distributions at or exceeding expectations, most of the goals for the computational pain quantification method were met. It was reasonably accurate for a remote monitoring context, the method proved robust in allowing

additional add-ons for possible future work, and the method worked with a large amount

of real data. This real data had a rather wide variety in image quality and lighting

conditions, meeting another criterion. Furthermore, it dealt with facial images and was

entirely automatic, and once trained, computation of a new pain intensity value could be

done in real time. Finally, Table 3.2 shows a breakdown of the current facial recognition

and detection methods, while Table 3.3 shows a breakdown of the computational pain

quantification method in comparison to existing methods based on accuracy and

classification methods.

Source	Method	Detection	Accuracy	Images
Chai 1999	Skin Color	Facial Detection	82%	80
Turk 1991	Eigenface	Facial Recognition	96%	2500
Liu 2000	Evolutionary Pursuit	Facial Recognition	92%	1107

Table 3.2: Comparison of Existing Methods to New Method

Source	Method	Detection	Accuracy	Images
Lo 2001	Eigenface	Emotion Recognition	70-80%	63
Smith 2001	FACS	Emotion Recognition	97%	451
Prkachin 2009	FACS	Pain Detection	46-88%	129
Monwar 2006	Eigenface	Pain Detection	86-93%	38 Videos
Monwar 2009	SVM	Pain Detection	96%	68 Videos
Gholami 2010	RVM	Pain Detection	91%	204
Ostberg 2014	Eigenface	Pain Quantification	84%	444

Table 3.3: Comparison of Existing Methods to New Method

The biggest concern at this point was that the method had only been tested on six

individuals, three if only considering those with large individual databases. It was

decided that at this point, the strength of personalized databases had been proven fairly

abundantly, due to the presented results and the nature of the eigenface method indicated

in the background material. However, what was needed was more raw data to validate the current pain quantification method.

CHAPTER 4 VALIDATION AND ADDITIONAL FIELDS

In order to validate the computational pain quantification method, it was decided that a significant amount of additional data was required. The robustness and personalized database aspects had already been proven sufficiently, so what was truly needed was lots of raw data from as many individuals as possible in order to validate the method. However, as it has been shown that group training databases are not as accurate as personalized training databases, a slightly more lax mean absolute error of 2 was chosen to validate the method. It is also required that the output data still be from the same distribution as the input data, although again, the p values looked for would be under less constraints due to group databases.

Furthermore, since additional data was needed, it was decided to determine if any additional fields would influence computational pain quantification. The four fields selected were age, gender, culture and cause. While most of these additional fields have been researched, most of the research has looked at one of two things; either the effects these fields has on pain tolerance [67], such as men having higher pain tolerances than women, or the effects these fields have on recognition [68], such as culture not being a distinguishing factor in human beings recognizing pain expressions. Nothing, however, looks at if any of these fields affect automatic pain quantification. For example, while pain might be recognizable cross culturally, it might be more difficult to determine pain levels from facial expressions from one culture verses another. As such, with the new data these four fields would be included and investigated as to their effect on automatic pain quantification.

In order to determine if these fields have any effect on automatic pain detection, the metrics for determining this will be relative accuracy. Assuming that the model is validated, the purpose of these additional fields will not be to look for mean absolute error, but instead to determine if the error is decreased by inclusion of these additional fields and to determine if the model is bias against these fields or certain aspects of these fields. The distributions will also be investigated to verify that they come from the same distribution.

4.1 Cross Sectional Study

The Cross Sectional Study was how the new data was acquired. Funded by the IBCRF and cleared by Marquette University's IRB and respective location specific review boards, the study collaborated with doctors from three different sites, Bangladesh, Nepal and South Dakota. It collected facial pain images, pain levels, the four fields discussed to be investigated, and a variety of other fields such as tiredness, from willing participants. The study lasted a year, starting in December 2012 and ending in December 2013. Over the course of this time period, 518 useable facial pain images with respective data were collected from distinct individuals. Tables 4.1 through 4.4 give the breakdowns of the number of images for each associated field and their corresponding data. Note that in most cases the totals do not add up to the 518 useable facial pain images. This is due to the fact that, as was their right as willing participants, some participants did not disclose all their information. While some of this information could have been estimated from their images, such as gender or age, it was deemed unethical to do so, as the participants did not provide the appropriate or necessary information and as such would be a violation of their rights.

Age Range	Number of Images
30 and Under	53
31-40	40
41-50	82
51-60	97
61-70	93
71 and Over	60

Table 4.1: Ages and Image Numbers

Cancer Type	Number of Images	Cancer Type	Number of Images
Bladder	12	Bone	3
Brain	5	Breast	65
Choriocarcinoma	1	Cervical	32
Colon and Rectal	50	Esophageal	7
Fibrosarcoma	2	Gallbladder	7
Germ Cell	2	Glottis	1
Leukaemia	2	Liver	11
Lung	79	Lymphoma	33
Mouth and Oropharyngeal	52	Multiple Myeloma	7
Omentum	1	Ovarian	25
Pancreatic	6	Periampullary	2
Pharynx	8	Plasmacytoma	1
Prostate	11	Renal	5
Sarcoma	9	Seminoma	3
Skin	10	Stomach	43
Testicular	2	Thymoma	2
Thyroid	5	Tongue	2
Uterus	3		

Table 4.2: Cancer Types and Image Numbers

Culture	Number of Images
Bangladesh	131
Nepal	311
South Dakota	71

Table 4.3: Cultures and Image Numbers

Gender	Number of Images
Female	256
Male	239

Table 4.4: Genders and Image Numbers

Of particular note is the Cancer Type table. The vast majority of the cancers listed contain only a handful of images. As the model is dependent on the number of images used, it was decided that for this field, only subfields that had 30 images or more would be included, resulting in the use of Breast, Cervical, Colon and Rectal, Lung, Lymphoma, Mouth and Oropharyngeal and Stomach Cancers as the subfields used.

4.2 Validation of Computational Pain Quantification Method

Originally, the Cross Sectional Study data was divided into two sets for validation purposes, training and testing. The training set contained 70 images and the testing set contained the remaining images. The model gave a mean absolute error of 3, which was completely unacceptable. However, due to the fact that eigenfaces can encode extraneous information such as lighting, glasses, facial hair, etc [66] and the data from the longitudinal study showed that individualized training sets improve performance, it was decided that this could be addressed. Since each image in the Cross Sectional Study is of a distinct individual, individualized databases were not an option. However, the

Longitudinal Study had a large database of only six individuals. As such, it was decided

to use the entire Longitudinal Study set as a training set and test the new Cross Sectional

Study data against it.

This resulted in a mean absolute error of 1.96, which was a significant

improvement, was considered satisfactory, and furthermore passed the Mann-Whitney U

test, indicating that it originates from the same distribution. The input and output graph is

shown in Figure 4.1

Figure 4.1: Input and Output Distributions for Cross Sectional Study Data

4.3 Additional Fields

For each additional field the subfields were divided into two categories, training and test. In order to keep the analysis to being strictly about the fields, each subfield had the same amount of training images, while the test image sets varied in number. Table 4.5 breaks down the training image sets sizes for each field.

Field	Training Set Size
Age	6
Cause	6
Culture	10
Gender	36

Table 4.5: Training Set Sizes by Field

Each field had one additional training set, a combined training set. This training set took random entries from all the other training sets within that field and would be used in order to determine relative accuracy. As this training does not discriminate based on any of the fields, each subfield test set would be compared against their own subfield training set and the combined field training set to get both the relative error and distribution. As the goal of this analysis is to determine if the method is biased towards any of these fields and if these fields affect the accuracy of the method, the difference vectors will be used to compute the error. That is, for each subfield, the relative error will be computed between the difference vectors brought about by testing with a subfield training set and the combined training set, taking the combined training set as the true values. Furthermore, as these are all vectors, p-norms will be used in place of absolute values with p being the length of the difference vectors.

4.3.1 Age

From the tables above, there were 6 subfields for the Age field, and each one had

6 training images. The relative errors and results of the Mann-Whitney U test are

presented below in Table 4.6. Note that an h value of 0 represents passing the test, and p

is the probability of observing the result if the medians are equal (test passed). As stated

in the previous chapter, the Mann-Whitney U test's null hypothesis is that the two

distributions have the same median, and is one method of determining if data comes from

the same distribution or not. As such, in this scenario, a test passed is indicative of both

difference vectors coming from the same distribution. These tests were all performed at a

95% confidence level, so while the chance is small; it is possible to incorrectly identify

distributions with this test.

Age Range	Training	Test	Relative Error	h	p
30 and Under	6	47	1.0432	0	0.0685
31-40	6	34	1.0443	1	0.0361
41-50	6	76	1.6667	1	0.003
51-60	6	91	1.3333	1	0.0284
61-70	6	87	4	0	0.7516
71 and Over	6	54	4	0	0.9847

Table 4.6: Relative Error for Ages

Of particular interest here is the apparent trend in the relative errors. For ages

under 40, the relative error is quite small, even if the difference distributions do not

always match. Even in the 41-60 age range the relative errors are still small and easily

tolerable, particularly as they have small p values. However, beyond that the relative

error blows up. This trend is indicative of automatic pain quantification becoming more

difficult as age increases, which in turn would mean that age should be a field included in future work to aid the algorithm's accuracy.

4.3.2 Cause

From the tables above, there were 7 subfields for the Cause field, and each one had 6 training images. The relative errors and results of the Mann-Whitney U test are presented below in Table 4.7.

Cancer Type	Training	Test	Relative Error	h	p
Breast	6	59	5	0	0.0871
Cervical	6	26	0.6667	0	0.9141
Colon and Rectum	6	44	5	0	0.2268
Lung	6	73	7	1	0.0051
Lymphoma	6	27	1	0	0.08
Mouth and Oropharyngeal	6	46	0.6667	0	0.899
Stomach	6	37	4	0	0.1149

Table 4.7: Relative Error for Causes

These relative errors appear to be rather varied, with some being very small and others being quite large. However, all but the largest relative error, Lung Cancer, passes the Mann-Whitney U test, indicating that the differences at least appear to be within the same distribution, although much like their errors, the p values range from very small to very large. This field may be worth future investigation based on these results, particularly if the cancer type could be even loosely correlated to location, one of the dimensions of pain. On the other hand, this data is indicative of the current method having not bias towards cancer type, which in turn indicates that it the method could be applicable to other chronic pain conditions and does not only apply to cancer.

4.3.3 Culture

From the tables above there were 3 subfields for the Culture field, and each one had 10 training images. The relative errors and results of the Mann-Whitney U test are presented below in Table 4.8

Culture	Training	Test	Relative Error	h	p
Bangladesh	10	121	1.0135	1	1.39E-19
Nepal	10	301	9	0	0.3348
South Dakota	10	61	2.5	1	9.79E-06

Table 4.8: Relative Errors for Cultures

The results of this field are rather interesting. While Bangladesh and South Dakota both have small relative errors, they fail the distribution test, albeit with tiny p values. Nepal however, has the highest relative error seen so far, yet passes the Mann-Whitney U test with a solid p value. Drawing conclusions from this is difficult, but it might be explained by over exaggeration of pain. It was encountered in the Longitudinal Study [13], where patients would over exaggerate their pain levels in order to get more attention from their health care provider. While this was observed and dealt with in Bangladesh during the Longitudinal Study, it might have been avoided due to collaborating with the same medical personnel. It might have occurred in Nepal, however, which over exaggeration of pain values could possibly explain this occurrence. This field warrants further study, particularly with why Nepal's relative error is so high.

4.3.4 Gender

From the tables above there were 2 subfields for the Gender field, and each one had 36 training images. The relative errors and results of the Mann-Whitney U test are presented in Table 4.9.

Gender	Training	Test	Relative Error	h	p
Female	36	220	6.0302	1	0.0211
Male	36	203	6.0329	0	0.2962

Table 4.9: Relative Errors for Gender

This field is by far the clearest to draw conclusions from. Despite the difference distributions being off, the relative errors are almost identical and rather high. Clearly the model could benefit from taking gender into account.

CHAPTER 5 CONCLUSION

5.1 Summary of Thesis

An automatic pain quantification tool could be extremely useful to medical personnel. Children below the ages of 3 or 4 are generally unable to describe their pain with any degree of accuracy [69]. Furthermore, a large percentage of elderly patients in nursing homes suffer from dementia, which can affect self-reporting of pain levels. A method able to determine pain levels for these patients would benefit both medical personnel and patients, helping doctors to prescribe the correct pain management techniques which in turn will aid the patient's quality of life. While the automatic pain quantification method developed in this thesis is designed with a remote monitoring context in mind, the end goal would be something to help patients in a hospital setting. This goal may not be as far off as it may seem, either, as it has been suggested that dividing the 0-10 pain level scale into segments of several pain levels could be quiet useful for some purposes [70]. Seeing as the current method is accurate within 1 pain level, this method should easily be able to provide a real time, highly accurate pain level if classified as low, medium and high pain, particularly as each level would span 3 pain levels. This would also be applicable for a remote monitoring context.

5.2 Contributions of Thesis

This thesis has provided three unique contributions. First and foremost, it has developed and validated a method for automatic pain quantification for application in a remote monitoring context. It is accurate when using an individual training set to within 1

unit of error, and within 2 when using a group training set. This thesis has also shown the effects of age, cause, culture and gender upon automatic pain quantification. While culture bears further investigation, age, cause and gender all have favorable results. Cause does not appear to be affected by the current method, which opens up the possibility of other chronic pain conditions, and that this method, while developed on cancer patients, should be able to be applied to other types of patients. Gender affects the method almost equally for both genders, while age affects the method more as it increases. Finally, two databases, totally almost a thousand pain images, have been created for further study. One of which also has the four fields discussed previously associated with each pain image, as well as others which have yet to be investigated, such as tiredness.

5.3 Impacts of Work

This work's most obvious impact will be on the quality of life for patients able to utilize this method. As this method was designed with a remote monitoring context in mind, patients will be able to ensure they are receiving adequate care from their doctors while maintaining their daily lives. It will also allow doctors to receive daily, or more frequent, pain data on their patients, allowing them to more accurately adjust pain medication, particularly as pain medication adjustments based on frequently determined pain levels improves patient outcomes and quality of life [10].

Furthermore, as this is a system designed with mobile health in mind, it is a system that could easily be adapted to a world where the majority of people have access to a cell phone. Not all of these phones are capable of sending photographs, but it is

obvious that cellphone ownership is a growing, global trend, so it is not unreasonable to assume that access to phones with that capability is also growing. Thus, this method could potentially provide pain quantification to people in developing countries.

Long term, this method will hopefully be able to be adapted to a real time, pain monitoring system that could potentially be put into hospitals and intensive care units to give doctors real time pain updates, which again would allow for even faster medication adjustments.

5.4 Future Work

In continuation of this work, several items could be investigated. The most obvious are the inclusion of age and gender within the algorithm, both have obvious effects on the accuracy of the method, so inclusion would improve the accuracy. For gender, this could be as simple as splitting training databases to only include one gender or the other, or it could be more complicated, but their inclusion will only help the method. Secondly, culture and cause should be investigated further. Nepal should be investigated as to whether or not over exaggeration of pain levels is an issue there, and cause should most definitely be investigated further. If the cause field presented here can be linked to location, location of pain would also be a useful inclusion in the method. Otherwise, causes other than cancer should also be included for study. Also, several other fields are available for further study that were collected by the Cross Sectional Study, and could warrant investigation. In addition, the Mann-Whitney U test was used exclusively in this thesis. The instances where it was used should be investigated for normality, and then use the stronger t test to give additional strength to these findings. Also, reducing

the number of eigenvectors can increase accuracy, by eliminating the 'noise' vectors that encode extraneous information. Removing these could potentially further improve accuracy. Finally, any way of taking this method and putting it in the hands of medical personnel should be looked at, whether it be simple remote monitoring, which could easily be done with the current method, or as a possible real time pain monitor. Anything that could possibly aid medical personnel in reducing the pain levels of their patients should be investigated in order to help improve the quality of life of everyone in pain.

BIBLIOGRAPHY

[1]   M. P. Jensen and P. Karoly, "Self-Report Scales and Procedures for Assessing Pain in Adults," in *Handbook Of Pain Assessment*, Guilford Press, 1992, pp. 15-75.

[2]   "AAPM Facts and Figures on Pain," the American Academmy of Pain Medicine, [Online]. Available: http://www.painmed.org/patientcenter/facts_on_pain.aspx. [Accessed 2 February 2014].

[3]   A. D. Woolf and B. Pfleger, "Burden of major musculoskeletal conditions," *Bull World Health Organ.,* vol. 81, no. 9, pp. 646-656, 2003.

[4]   W. F. Stewart, J. A. Ricci, S. R. Hahn and D. Morganstein, "Cost of lost productive work time among US workers with depression.," *JAMA,* vol. 289, no. 23, pp. 3134-3144, 2003.

[5]   P. Mantyselka, E. Kumpusalo, R. Ahonen and et al., "Pain as a reason to visit the doctor: a study in Finnish primary health care," *Pain,* vol. 89, no. 2-3, pp. 175-180, 2001.

[6]   R. Chapman, "New JCAHO Standards for Pain Management: Carpe Diem!," *APS Bulletin,* vol. 2008, 2000.

[7]   J. Stephenson, "Veterans' pain a vital sign," *JAMA,* vol. 281, no. 11, p. 978, 1999.

[8]   R. S. Salcido, "Is pain a vital sign?," *Adv Skin Wound Care,* vol. 16, no. 5, p. 214, 2003.

[9]   D. B. Gordon, J. L. Dahl, C. Miaskowski and et al, "American pain socity recommendations for improving the quality of acute and cancer pain management: American Pain Socity Quality of Care Task Force," *Arch Intern Med.,* vol. 165, no. 14, pp. 1574-1580, 2005.

[10] P. Lucey, J. Cohn, K. Prkachin, P. Solomon, S. Chen and I. Mathews, "Painful monitoring: Automatic pain monitoring using the UNBC-McMaster shoulder pain expression archive database," *Image and Vision Computing,* pp. 197-205, 2012.

[11] D. West, "How mobile devices are transforming healthcare," May 2012. [Online]. Available: http://www.brookings.edu/research/papers/2012/05/22-mobile-health-west.

[12] The Center for Children with Special Health Needs Children's Hospital and Regional Medical Center, Seattle, WA, June 2006. [Online]. Available: http://www.nwsicklecell.org/uploads/file/Sickle%20Elements%20of%20Care.pdf.

[Accessed February 2014].

[13] M. M. Haque, "Mobile Based Sympton Management for Palliative Care," May 2013. [Online]. Available: http://epublications.marquette.edu/cgi/viewcontent.cgi?article=1267&context=disse rtations_mu. [Accessed February 2014].

[14] United Nations, "Deputy UN chif calls for urgent action to tackle global sanitation crisis," UN News Centre, 21 March 2013. [Online]. Available: http://www.un.org/apps/news/story.asp?NewsID=44452&Cr=sanitation&Cr1=#.Uvb CvbQhirr. [Accessed February 2014].

[15] P. K. Adelmann and R. B. Zajonc, "Facial efference and the experience of emotion," *Annual Review of Psychology,* no. 40, pp. 249-280, 1989.

[16] A. Mehrabian, "Communication without Words," *Psychology Today,* vol. 2, no. 4, 1968.

[17] K. D. Cragi, K. M. Prkachin and R. E. Grunau, "The Facial Expression of Pain," in *Handbook of Pain Assessment*, Guilford Press, 1992, pp. 153-169.

[18] D. Chai and K. N. Ngan, "Face Segmentation Using Skin-Colo Map in Videophone Applications," *IEEE Transactions on Circuits and Systems for Video Technology,* vol. 9, no. 4, pp. 551-564, 1999.

[19] D. Gokalp and S. Aksoy, "Finding Faces in News Videos," in *4th International Workshop on Content-Based Multimedia Indexing*, Riga, Latvia, 2005.

[20] S. K. Singh, D. S. Chauhan, M. Vatsa and R. Singh, "A Robust Skin Color Based Face Detection Algorithm," *Tamkang Journal of Science and Engineering,* vol. 6, no. 4, pp. 227-234, 2003.

[21] L. I. Smith, "A tutorial on Principal Components Analysis," 26 February 2002. [Online]. Available: http://www.cs.otago.ac.nz/cosc453/student_tutorials/principal_components.pdf. [Accessed February 2014].

[22] M. Turk and A. Pentland, "Eigenfaces for Recognition," *Journal of Cognitive Neuroscience,* vol. 3, no. 1, pp. 72-86, 1991.

[23] W. S. Yambor, B. A. Draper and J. R. Beveridge, 1 July 2000. [Online]. Available: http://www.cs.colostate.edu/evalfacerec/papers/eemcvcsu.pdf. [Accessed February 2014].

[24] V. Akalin, *Face Recognition Using Eigenfaces and Neural Networks,* 2003.

[25] M. Rizon, M. F. Hashim, P. Saad, S. Yaacob, M. R. Mamat, A. Y. Shakaff, A. R. Saad, H. Desa and M. Karthigayan, "Face Recognition using Eigenfaces and Neural Networks," *American Journal of Applied Sciences,* vol. 2, no. 6, pp. 1872-1875, 2006.

[26] M. Agarwal, H. Agrawal, N. Jain and M. Kumar, "Face Recognition using Principle Component Analysis, Eigenface and Neural Network," 2010.

[27] A. Pentland, B. Moghaddam and T. Starner, "View-Based and Modular Eigenspaces for Face Recognition," in *Computer Vision and Pattern Recognition*, Seattle, WA, 1994.

[28] T. Heseltine, N. Pears and J. Austin, "Evaluation of image pre-processing techniques for eigenface based face recognition," in *The Proceedings of the Second International Conference on Image and Graphics*, 2002.

[29] C. Liu and H. Wechsler, "Evolutionary Pursuit and Its Application to Face Recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 22, no. 6, pp. 570-582, 2000.

[30] P. Ekman and W. Friesen, "Manual for the Facial Action Codying System," *Palo Alto: Consulting Psychologists Press,* 1978.

[31] B. Braathen, M. Bartlett, G. Littleword-Ford, G. E. Smith and J. Movellan, "An approach to automatic recognition of spontaneous facial actions," in *Fifth International Conference on automatic face and gesture recognition*, 2002.

[32] M. Bartlett, G. Littlewort, J. Whitehill, E. Vural, T. Wu, K. Lee, A. Ercil, M. Cetin and J. Movellan, "Insights on spontaneous facial expressions from automatic expression measurement," *Dynamic Faces: Insights from Experiements and Computation,* 2006.

[33] M. Bartlett, G. Littlewort, C. Lainscsek, I. Fasel, M. Frank and J. Movellan, "Fully automatic facial action recognition in spontaneous behaviour," *7th International Conference on Automatic Face and Gesture Recognition,* 2006.

[34] E. Smith, M. Bartlett and J. Movellan, "Computer recogntion of facial actions: A study of coarticulation effects," in *8th Annual Joint Symposium on Neural Computation*, 2001.

[35] K. Prkachin and S. Mercer, "Pain expressions in patients with shoulder pathology: validity, coding properties and relation to sickness impact," *Pain,* vol. 39, pp. 257-65, 1989.

[36] G. Littlewort, M. Bartlett and K. Lee, "Faces of Pain: Automated measurement of spontaneous facial expressions of genuine and posed pain," San Diego, CA, 2006.

[37] P. S. Aleksic and K. K. Aggelos, "Automatic Facial Expression Recognition Using Facial

Animation Parameters and Multistream HMMs," *IEEE Transactions on Information Forensics and Security,* vol. 1, no. 1, pp. 3-11, 2006.

[38] L. Trujillo, G. Olague, R. Hammoud and B. Hernandez, "Automatic Feature Localization in Thermal Images for Facial Expression Recognition," *Computer Vision and Pattern Recognition - Workshops,* pp. 14-21, 2005.

[39] M. C. Su, Y. J. Hsieh and D. Y. Huang, "A Simple Approach to Facial Expression Recogntion," in *WSEAS International Conference on Computer Engineering and Applications*, Gold Coast, Australia, 2007.

[40] G. R. Murthy and R. S. Jadon, "Recognizing Facial Expressions using Eigenspaces," in *International Conference on Computational Intelligence and Multimedia Applications*, 2007.

[41] G. R. S. Murthy and R. S. Jadon, "Effectiveness of Eigenspaces for Facial Expressions Recognition," *International Journal of Computer Theory and Engineering,* vol. 1, no. 5, pp. 638-642, 2009.

[42] H. B. Deng, L. W. Jin, L. X. Zhen and J. C. Huang, "A New Facial Expression Recognition Method Based on Local Gabor Filter Bank and PCA plus LDA," *International Journal of Information Technology,* vol. 11, no. 11, 2005.

[43] C. Frank and E. Noth, "Automatic Pixel Selection for Optimizing Facial Expression Recognition using Eigenfaces," *DAGM,* pp. 378-385, 2003.

[44] C. Frank and E. Noth, "Optimizing Eigenfaces by Face Masks for Facial Expression Recognition," in *Computer Analysis of Images and Patterns*, Berlin, Springer Berlin Heidelberg, 2003, pp. 646-654.

[45] I. Eguchi and K. Kotani, "Facial Expression Analysis by Generalized Eigen-Space Method Based on Class-Feautres (GEMC)," in *Image Processing*, 2005.

[46] Y. Kosaka and K. Kotani, "Facial Expression Analysis by Kernel Eigenspace Method Based on Class Features (KEMC) Using Non-Linear Basis for Separation of Expression-Classes," in *International Conference on Image Processing*, 2004.

[47] D. T. Lin, "Facial Expression Classification Using PCA and Hierarchical Radial Basis Function Network," *Journal of Information Science and Engineering,* no. 22, pp. 1033-1046, 2006.

[48] H. C. Lo and R. Chung, "Facial Expression Recognition Approach for Performance Animation," in *Digital and Computational Video*, Tampa, FL, 2001.

[49] K. Craig, R. Grunau and J. Aquan-Assee, "Judgement of pain in newborns: facial activity and cry as determinants," *Can J Behav Sci,* vol. 20, pp. 422-51, 1988.

[50] K. Craig, H. Hadjistavropoulos, R. Grunau and M. Whitfield, "A comparison of two measures of facial activity during pain the newborn child," *Journal of Pediatric Psychology,* vol. 19, pp. 305-18, 1994.

[51] K. Craig, M. Whitfield, R. Grunau, J. Linton and H. Hadjistavropoulos, "Pain in the preterm neonate behavioural and physiological indices," *Pain,* vol. 52, pp. 287-99, 1993.

[52] R. Grunau and K. Craig, "Pain expression in neonates: facial action and cry," *Pain,* vol. 28, pp. 395-410, 1987.

[53] R. Grunau, C. Johnston and K. Craig, "Neonatal Facial and cry responses to invasive and non-invasive procedures," *Pain,* vol. 42, pp. 295-305, 1990.

[54] S. Brahnam, C. Chuang, F. Shih and M. Slack, "Machine recognition and representation of neonatal facial displays of acute pain," *Artif. Intel. Med.,* vol. 36, pp. 211-222, 2006.

[55] S. Brahnam, L. Nanni and R. Sexton, "Introduction to neonatal facial pain detection using common and advanced face classification techniques," *Stud. Comput. Intel.,* vol. 48, pp. 225-253, 2007.

[56] K. Prkachin and P. Solomon, "The structure, reliability and validity of pain expression: Evidence from patients with shoulder pain," *Pain,* vol. 139, pp. 267-274, 2009.

[57] A. Ashraf, S. Lucey, J. Cohn, T. Chen, Z. Ambadar and K. Prkachin, "The painful face II- Pain expression recognition using active appearance models," *Image and Vision Computing,* vol. 27, pp. 1788-1796, 2009.

[58] G. Littlewort, M. Bartlett and K. Lee, "Automatic Coding of Facial Expressions Displayed During Posed and Genuine Pain," *Image and Vision Computing,* vol. 12, no. 27, pp. 1741-1844, 2009.

[59] M. M. Monwar and S. Rezaei, "Pain Recognition Using Artificial Neural Network," in *IEEE International Symposium on Signal Processing and Information Technology*, 2006.

[60] M. M. Monwar and S. Rezaei, "Video Analysis for View-Based Painful Expression Recognition," in *International Joint Conference on Neural Networks*, 2008.

[61] M. M. Monwar and S. Rezaei, "Appearance-based Pain Recognition from Video Sequences," in *International Joint Conference on Neural Networks*, Vancouver, BC, Canada, 2006.

[62] M. M. Monwar, S. Rezaei and K. Prkachin, "Eigenimage Based Pain Expression Recognition," *International Journal of Applied Mathematics,* vol. 36, no. 2, 2007.

[63] M. M. Monwar and S. Rezaei, "Support vector machine for automatic pain recognition," *Computational Imaging VII,* 2009.

[64] B. Gholami, W. Haddad and A. Tannenbaum, "Relevance Vector Machine Learning for Nenate Pain Intensity Assessment using Digital Imaging," *IEEE Trans. Biomend. Eng.,* 2010.

[65] Y. Ayzenberg, J. Hernandez and R. W. Picard, "FEEL: Frequent EDA and Event Logging - A Mobile Social Interaction Stress Monitoring System," in *CHI*, 2012.

[66] W. Yambor, B. A. Draper and J. R. Beveridge, "Analyzing PCA-based Face Recognition Algorithms: Eigenvector Selection and Distance Measures," in *Second Workshop on Empirical Evaluation Methods in Computer Visio*, 2000.

[67] K. W. Woodrow, G. D. Friedman, A. B. Siegelaub and M. F. Collen, "Pain Tolerance: Differences According to Age, Sex and Race," *Psychosomatic Medicine,* vol. 34, no. 6, pp. 548-556, 1972.

[68] P. Ekman, E. R. Sorenson and W. V. Friesen, "Pan-Cultural Elements in Facial Displays of Emotion," *Science,* vol. 164, no. 3875, pp. 86-88, 1969.

[69] R. A. Kozol and A. Voytovtch, "Misinterpretation of the Fifth Vital Sign," *Archives of Surgery,* vol. 142, no. 5, pp. 417-419, 2007.

[70] N. J. Seo, B. S. Sindhu and O. Shechtman, "Influence of Pain Associated with Musculoskeletal Disorders on Grip Force Timing," *Journal of Hand Therapy,* pp. 335-344, 2011.