January 2011

Digital Image Analysis Of Gram Stained Culture Specimens

Ronald George Hauser
Yale School of Medicine, hauserrg@gmail.com

Digital Image Analysis of Gram Stained Culture Specimens

A Thesis Submitted to the

Yale University School of Medicine

in Partial Fulfillment of the Requirements for the

Degree of Doctor of Medicine

by

Ronald George Hauser III

2011

DIGITAL IMAGE ANALYSIS OF GRAM STAINED CULTURE SPECIMENS

R. George Hauser III and Sheldon M. Campbell. Section of Microbiology, Department of Laboratory Medicine, Yale University, School of Medicine, New Haven, CT.

Digital image analysis for the interpretation of images in clinical microbiology has many potential advantages over current practices. Compared to traditional image interpretation by a medical technologist, digital image analysis offers standardization between laboratories, round-the-clock interpretation, and quantitative results. In the first study of its kind known to the authors, a digital image analysis program was prototyped to interpret a slide containing Gram stained microorganisms. The sample microorganisms were obtained from culture plates during routine processing and subjected to Gram's stain. An initial study learned from 11 Gram-stained slides and classified their microorganisms into the group: Gram-positive, Gram-negative, rods, coccus, and yeast. The sensitivity of identification ranged from 66% to 99% and the specificity ranged from 78% to 99%. The algorithm was next applied to a larger set of 78 slides. The accuracy rate for slide classification was 60 out of 78 or 77%. After using this larger dataset to train the algorithm, the accuracy rate for individual objects was 94% averaged over 5 trials. This suggests the parameters used by the algorithm can differentiate between groups, and the lack of accuracy in classifying the larger database occurred due to limitations in the original training data. Overall, the project demonstrates a unique application of digital image analysis to clinical microbiology.

Acknowledgments

I would like to take a moment to thank those people who contributed to this thesis. Dr. Sheldon Campbell provided unspoken inspiration by believing this project could come to fruition, and he had great patience waiting for the results. The microbiology staff at the West Haven Veterans Affairs Hospital took the time to share their samples to create the images used for analysis. My parents, Ron and Christine Hauser, provided their overall support and shared technical expertise in brainstorming mechanical solutions to future automation. Caitlin Koerber made an intellectual contribution to the project. Finally, thank you Christine for everything.

Table of Contents

Introduction

Digital image analysis is an established technique approved by the U.S. Food and Drug Administration for application to medical areas as diverse as Pap smear tissue screening and radiologic mammography interpretation. To provide a general example, it involves digitizing an image either from a microscope or radiograph followed by interpretation using an algorithm to draw inferences about the image. In clinical microbiology, the microscope has multiple uses through all stages of workflow, each providing a potential application for digital image analysis. Digital image analysis has advantages over most new technology because it has no additional consumable costs and relies on mass-produced capital like computes and color sensors. Compared to the current interpretation by trained specialists, it could operate nights and weekends, support untrained or non-specialist personnel, provide quantitative results, and enhance standardization across institutions. At the present time, a shortage in the supply of trained medical technologists exists compared to the demand, which is predicted by the American Society of Clinical Pathologists to develop into a public health crisis in the next 5 to 10 years. As the cost of labor increases because demand for skilled labor further outweighs the supply, increased automation becomes more economically feasible. Substantial automation of the current microbiology lab will necessarily involve digital image analysis because of the fundamental role played by microscopy in specimen processing, initial assessment, and continued workup of positive cultures (1). Alternative technologies, including mass spectrometry and nucleic acid amplification

may change clinical microbiology, eventually making the microscope and digital image analysis obsolete. But, in the near future these technologies seem unlikely to outcompete the microscope's speed and cost in all of its versatile applications. Consequently, projects in digital image analysis of microscope images will have useful applications in the clinical microbiology laboratory.

Microscopy plays an essential role in the processing of clinical microbiology specimens (2). It is the first step in the processing of most specimens, guiding the selection of appropriate isolation media and the initiation of empirical antibiotic therapy (2). Microscopy may also be used as an intermediate step in processing. For instance, the microscopic interpretation of the Gram stained specimen is used to select the appropriate plate to inoculate for the Vitek2 biochemical identification system (3). Final identification can also involve microscopy, particularly for fungi and parasites. Overall microscopy is a fundamental technique of clinical microbiology (2).

Microscopy has widespread use in clinical microbiology, but the requirement for highly trained personnel for interpretation is a major drawback because hiring difficulties have existed for at least the past 20 years. The vacancy rates for the staff-level medical technologist was 13.8% in 1992(4), 12.5% in 2000(5), and 10.4% in 2009(6). Although the labor vacancy rate has remained historically high, it is also the anticipated increase in the absolute number of vacancies that has lead to the prediction of a healthcare crisis by the American Society of Clinical Pathology. As explained by the United States Bureau of Labor Statistics in 2007, an additional 149,000 medical

technologists and technicians will be needed by 2014 to replace the retirees of the baby boom generation and meet their health care demands(7).

The scarcity of trained personnel has particularly affected microbiology because, unlike clinical chemistry or hematology, it relies heavily on manual labor (1). As a consequence of the increasing ratio of technologist labor demand to supply, the cost of available labor will presumably increase. Thus, the relative scarcity of skilled labor in clinical microbiology and the predicted increases in labor costs motivate alternative solutions for microscopic interpretation.

Digital image analysis is a potential adjunct or alternative to interpretation by a microbiology technologist. The technology is not new. In 1985 the U.S. Food and Drug Administration (FDA) approved an application to interpret antibiotic susceptibility discs using digital image analysis(8). Since then it has permeated multiple specialties including anatomic pathology and radiology (9-11).

Giles Scientific Inc. received approval from the FDA for antibiotics susceptibility disc interpretation using digital image analysis in 1985. Their current systems, the BiomicV3 and TrinityV3, can read antibiotic susceptibilities from multiple sources including disc diffusion, D-test and the Etest. Since their initial device capable of interpreting antibiotic susceptibility, they have expanded to organism identification and colony counting. The system identifies organisms by interpreting biochemical panels and chromogenic agar(12). Their system is likely one of the first applications of digital image analysis to clinical pathology.

Technology for image interpretation is also used in anatomic pathology. For example, the decision to treat a breast cancer patient with Herceptin (Trastuzumab) depends on the quantified level of Human Epidermal growth factor Rector 2 (HER2). Digital image processing programs created by Aperio and another company, BioImagene, were approved by the U.S. Food and Drug Administration to aid a qualified pathologist in quantifying the immunohistochemical (IHC) stain for HER2 (11). The interpretation of these stains was quite variable between pathologist, and the use of digital image analysis decreases inter-observer variability.

Another implementation of the technology is used in the screening of the Pap smear. NeoPath's AutoPap System is approved by the U.S. Food and Drug Administration for automated screening of Pap smears used to detect cervical cancer (10). In 1999 a prospective study using over 25,000 patient slides from 5 clinical laboratories, the AutoPap System outperformed the current practice with reduced false-negative and false-positive results (10).

In radiology, computer-aided detection (CADe) and computer-aided diagnosis (CADx) algorithms exist to aid the physician in detecting the presence of disease (9). To validate and compare these algorithms, large databases of mammography images have been created by various institutions including the Digital Database for Screening Mammography (Massachusetts General Hospital, University of South Florida), the Mammographic Image Analysis Society's Mammographic Database (United Kingdom), and B-Screen (Netherlands) among others(13).

Digital image analysis has demonstrated its usefulness across multiple medical disciplines, and it has numerous potential applications in clinical microbiology. To begin, the technology could be incorporated into blood culture monitoring systems to automatically report positively-screened blood cultures. A compact version of the system could be amenable to clinic use as a point of care test. As a portion of a larger automated system, it could speed the flow of specimens through the microbiology lab by triaging positive culture plates. Digital image analysis applications exist where the microscope is found, namely across a broad array of clinic and lab settings.

In general, automation has arrived slowly to the clinical microbiology lab compared to other areas of laboratory medicine. For example much of the work in clinical chemistry is fully automated(1). Prior attempts at automation of microbiology have identified tasks involving visual recognition and manual dexterity as the primary impediments (1). However, in the past decade, visual recognition technology has become relatively inexpensive, as color sensors and hardware for processing images exist at reasonable prices (1). Additionally the cost for robotics capable of performing dexterous tasks continues to decrease (14). As automation technology becomes inexpensive in absolute dollars and relative to the cost of labor, automation in clinical microbiology becomes cost-effective.

While digital image analysis may fit within the current structure of the microbiology lab, other techniques have the theoretical ability to change the structure of the microbiology lab altogether, making the niche for digital image analysis obsolete.

A brief examination of these technologies is warranted before developing an application in digital image analysis for clinical microbiology. Two of these developments are mass spectrometry and nucleic acid-based technologies.

Mass spectrometry has many flavors, and only a few are amenable to bacterial identification (15). Matrix-assisted laser desorption/ionization (MALDI) time-of-flight (TOF; hence, MALDI-TOF) mass spectrometry, unlike traditional mass spectrometry allows the ionization of large biomolecules required for bacterial identification. Traditional mass spectrometry destroys these large biomolecules during the ionization step resulting in unpredictable analytic results, a phenomenon termed “harsh ionization”. MALDI-TOF mass spectrometry with its “soft ionization” step aerosolizes macromolecules without disrupting their structure. The majority of these macromolecules are abundant ribosomal bacterial proteins, but other macromolecules are included as well (15). The same isolate can produce different spectra depending on the growth media, growth state, and sample preparation (15). Presumably the macromolecules detected by MALDI-TOF mass spectrometry are differentially expressed under various conditions.

Despite this limitation, MALDI-TOF mass spectrometry system has the potential to replace the traditional biochemical profiles for bacterial identification. Most isolates can be identified easily under growth conditions common in the clinical laboratory. The system can usually operate on a single bacterial colony, but colonial growth typically takes 12-48 hours to occur, an inherent limitation of the technology at this time. (15).

In addition, colony isolation is required to produce an analyzable spectrum, which adds to the labor required for interpretation. Chemical treatment to disrupt the cell wall or trypsin-aided disruption of proteins may also be required to increase the ability to distinguish similar organisms (15). Because it requires traditional growth and isolation of multiple organisms, it seems more likely to replace the current biochemical profiles for bacterial identification than fill the role of workflow triage performed by traditional microscopy.

Nucleic acid amplification methods have become widely used in clinical microbiology, but are still primarily tools for detecting single, specific pathogens, as opposed to assessing the microbial flora of a primary specimen. Multiplex PCR methods are becoming more widely-available, but still fail to detect the wide range of bacterial and fungal pathogens of the Gram stain, though the molecular methods provide more specific taxonomic information.

In contrast to protein mass spectrometry, nucleic acid amplification followed by mass spectrometry analysis of nucleotide fragments is extremely sensitive and therefore does not require culture isolation or enrichment (15). Its primary disadvantage is cost, typically 10 to 100 times more expensive than mass spectrometry of culture-amplified material (15). It also requires an additional DNA purification step, and is in a comparatively early stage of development with significant unknowns. While mass spectrometry of bacterial proteins or amplified nucleic acids may become the future, none have approval for use in diagnostic applications (15).

The continued reliance on the microscope to initiate empiric antibiotics and direction of microbiology workflow seems assured for the immediate future. Therefore, the development of digital image analysis techniques will have a potential home in the microbiology laboratory either as an adjunct to current methods or as a component of a larger automation initiative.

The digital interpretation of Gram-stained organisms from a culture plate is a reasonable first step to the introduction of digital image analysis to microbiology. The Gram stain is the most commonly performed differential stain in microbiology (2). It has the ability to distinguish major classes of bacteria and yeast. The decision to use organisms removed from a culture plate simplifies the problem relative to a more complex direct specimen sample because the specimen taken from a culture has less interfering background material. Furthermore, no literature currently exists on the classification of Gram-stained organisms using computational techniques. The closest relative to digital image analysis of Gram-stained microorganisms were a handful of articles from the sewage treatment literature that apply image analysis to organisms found in sewage treatment plants, but they do not attempt to identify organisms (14-16). Thus a project involving digital image analysis of Gram-stained organisms from culture was pursued.

In summary, the microscope is a critical tool in clinical microbiology, used at multiple steps in specimen workflow from initiation of empiric antibiotics to specimen triage and even final diagnoses. Limitations of microscopy include the relative scarcity

of specialized personnel capable of interpreting microscopic slides, increasing labor costs, the absence of off-hour interpretation, subjective interpretations, and incomplete standardization. Digital image analysis potentially requires a modest investment in capital, no additional consumable costs, and would be operational at all hours by a non-specialist to provide standardized, quantitative results. Digital image analysis is an established technique approved by the U.S. Food and Drug Administration for use in clinical microbiology to interpret biochemical panels and quantify antibiotic susceptibility. Although new technologies like mass spectrometry of bacterial proteins and nucleic acid amplification may eventually make the microscope obsolete, these technologies seem unlikely to compete with the microscope's low cost, rapid results, and diversity of applications in the near future. These systems appear more likely to replace conventional biochemical and other bacterial culture identification systems than to replace the microscope. Thus, an investigation into the use of digital image analysis for the interpretation of microbiology images represents a worthwhile effort.

Purpose

The purpose of this study is to investigate the feasibility of microorganism classification using digital image analysis.

Hypothesis

A prototype software program can correctly classify the microorganisms on a digital image into Yeast, Gram-positive cocci, Gram-positive rods, Gram-negative cocci, or Gram-negative rods.

Specific aims

- Obtain Gram-stained slides with microorganisms taken from culture plates.
- For each slide, capture representative images of the microorganisms contained within the slide.
- Prototype a software program with the ability to distinguish major microorganism groups through the interpretation of digital images.
- Test the ability of the software to correctly identify the microorganisms given a set of images captured from the slide.

Methods

The methods section is separated into subsections detailing the slide database, classification program, and experiment design. The set of slides used to capture images for the image processing experiments are referred to as the slide database.

Slide Database

A single set of Gram-stained slides were used to create two separate image repositories. The first image repository used fewer slides and has fewer images compared to the second image repository. It was created earlier in the project when the number of slides was limited. The second image repository contains pictures of all available slides. Experiment 1 uses the first image repository, while experiment 2 and 3 use the second image repository. A description of the slides found in the slide database is listed in Table 1.

The slides for the first experiment included 11 in total: 2 yeast, 4 Gram-positive cocci, no Gram-negative cocci, 1 Gram-positive rod, and 4 Gram-negative rods. Experiments 2 and 3 used 78 slides: 15 Gram-negative rods, no Gram-negative cocci, 46 Gram-positive cocci, 9 Gram-positive rods, and 8 yeasts. The number of images in the first image database was 22 or 2 from each slide. The second image database contains 780 images or 10 from each slide. Both databases contain true color images. The images have pixel counts of 1360 by 1024. They are stored in uncompressed tagged

image format, which is commonly known by the ".TIF" file extension. The total data contained in the second library is approximately 4.14 gigabytes.

Images for both image repositories were captured by an Olympus DP71 Microscope Digital Camera mounted on an Olympus BX51 Microscope. The slides were viewed with the 100x objective under oil immersion. The image fields were selected for clearly-demonstrated morphology and well-separated organisms on the slide. In general they avoided highly dense bacterial clusters found near the center of the slide, and instead focused on the periphery of the slide where isolated organisms were more likely to reside. To abate a theoretical selection bias from image collection, learning algorithms were prevented from measuring quantities relevant to the overall image. Instead they were restricted to analyzing sub-images. (The concept of sub-image is explained in the image analysis section.) As an example, images of Gram-negative rods may have many more organisms on a slide compared to yeast. The use of organism quantity per image may help in the classification of the slide into Gram negative rod or yeast, but could also occur as a confounding variable associated with the image capture method.

The slides were collected by the VA Hospital in West Haven, Connecticut from November 2010 to March 2011. They contained Gram-stained microorganisms transferred directly from culture plates to the glass slide. The slides were made during the routine processing of specimens, and saved for this project instead of being discarded. The slides came from all areas of the microbiology lab performing Gram

stains. The Gram stain was performed in the conventional way following the clinical laboratory's standard procedure (2). Medical identification numbers were included with each slide to facilitate recording of the final identified organism present on the slide. When possible the microorganism on the slide was paired with a species level identification. A portion of the slides did not receive species level classification. For these slides only the interpretation of the Gram-stained microorganisms by the medical technologist was included in the database.

Two slides were excluded from the study because their final identification did not match the organism type found on the slide. For example, the final identification on one slide listed the organism as *Staphylococcus aureus*, while the Gram stain contained yeast. Another slide was excluded from the study because it was damaged by repeated viewing under oil immersion. All other slides were included, including a number of over-decolorized slides. The inclusion of all available slides was meant to make the project more realistic to implementation in a typical microbiology lab.

Digital Image Analysis

The image analysis program is a type of learning algorithm. Analogous to human learning the program has two separate modes of operation: train and test. The training phase is similar to studying before a test. The program is provided a set of labeled data. One example in the labeled data would be an image of yeast and the label "yeast". The program studies the training data in the training phase to discover a classification rule. Next, the classification rule is put to the test. A set of data without labels is provided to

the program, and the program generates labels. For example, an image of yeast is given to the program and the program outputs “yeast”. With the learning analogy in mind, let us be more concrete about the program details.

The initiation of the training mode requires a set of sub-images with their corresponding labels. These sub-images are not the original pictures captured by the digital camera. Rather they are a processed version of the original picture. The image captured directly from the camera is separated into these sub-images, which may contain the microorganisms. The process of creating the sub-images from the original image is called segmentation. The input and output of segmentation is shown in Figure 1. As shown in Figure 1, the output of segmentation includes sub-images that may represent microorganisms, smudges on the microscope optics, or concentrations of crystal violet stain.

After the creation of the sub-images by segmenting the original image, a label is applied to each sub-image. The label is provided by a curator, who painstakingly views each individual sub-image and assigns a label. In the following experiments multiple labels were assigned to individual sub-images. These labels correspond to categories in the algorithms classification scheme. One label identified the shape: coccus, rod, “yeast-shape”, or undetermined. Another label identified the color: Gram-positive bacteria, “yeast color”, Gram-negative bacteria, or undetermined. In this labeling scheme yeast has a separate shape and color designation because both shape and color attributes distinguished yeast from bacteria. For example, the shape of yeast is typically

more round than a coccus because they are larger cells. Their color also distinguishes them from bacteria because the yeast nucleus typically retrains more crystal violet stain than the cytosol. To designate an organism as yeast it must meet both the shape and color criteria. At this point, the training data is complete because a set of sub-images exist with corresponding labels.

The training data is input into the learning algorithm to discover a classification rule. In the process of learning, the algorithm calculates multiple features for each of the sub-images. These features are similar to those used by microbiologists to classify microorganisms. Examples include color, size, and shape. At the conclusion of the training phase the algorithm has a method to classify any potential sub-image.

For this prototype program two separate learning algorithms were included. The first identified the shape of the microorganism: coccus, rod, “yeast-shape”, or undetermined. The other identified the color: Gram-positive bacteria, Gram-negative, “yeast-color”, or undetermined.

The test is the second phase of the learning algorithm. Similar to the training phase, a set of sub-images is segmented from the original image. Unlike the training phase, where each sub-images received a label for color and shape from a human curator, the sub-images used for the test did not. Instead the learning algorithm provided the label.

The evaluation of the learning algorithm occurred by comparing its classification for each sub-image to the classification label of the slide from where the sub-image originated. For instance, a slide of Gram-positive cocci had an image taken and segmented into sub-images. The correct label for all sub-images would be "Gram-positive" for the color classifier and "cocci" for the shape classifier.

Overview of Experiment Design

Three experiments are performed. Each experiment uses different training and evaluation methods to test the different aspects of the classification prototype.

The first experiment evaluates the classification program in identifying the traits of shape and color in microorganisms. It uses the smaller image database of 11 images. The images were segmented into sub-images. A portion of the total sub-images were used as training data and the remainder saved for a test. The training sub-images were labeled by a human curator. Two labels were used for each sub-image. A label was provided for color and a second label for shape. One learning algorithm was trained to recognize color and another to recognize shape. In the test phase the sub-images not used in training were classified by each learning algorithm. The results were tabulated independently for each color or shape classification category by comparing the result of the two classifiers to the original label of the slide.

The second experiment tests the classification program in assigning a label to an unknown slide. It uses the larger image database of 780 images from 78 slides. The

classifiers trained in the first experiment are used without additional training. For each slide, a set of 10 images are segmented and their sub-images classified. A classification label for the slide is derived by taking the mode of the separate categories for shape and color for the sub-images belonging to the slide.

The third experiment evaluates the capacity of the learning algorithm to classify the sub-images in the larger image database given additional training. The training dataset is a random selection of sub-images from the major categories: yeast, Gram-positive cocci, Gram-positive rods, and Gram-negative rods. The labels in the training dataset are the slide labels. The test data is the unlabeled training data. The learning algorithm is setup to prevent memorizing the training data. The result is the proportion of sub-images correctly labeled in the test. The test is repeated after selecting new training data.

The results of the experiments were compiled using standard validation techniques for learning algorithms. Sensitivity and specificity for each category is listed for experiment 1, which is a standard technique for evaluating performance of a learning algorithm (16-19). Confusion matrices are presented for experiments 2 and 3. This is also a standard method for validation of learning methods (20-23).

Results

Experiment 1:

In order to test the ability of the program to classify microorganisms, 2 representative images were captured from each of 11 different Gram-stained slides for a total database of 22 images. Next the background of the image was removed in the process of image segmentation. The nonbackground portion of the images formed 7,096 sub-images separated by 8-connectivity between pixels. The total composition of the image library was yeast (4 images = 635 sub-images), Gram+ cocci (8 images = 1012 sub-images), Gram- cocci (zero), Gram+ rods (2 images = 857 sub-images), and Gram- rods (8 images = 4592 sub-images). These sub-images consisted of all types of non-background objects: microorganisms, unidentifiable organism fragments, concentrations of crystal violet stain, and debris from the microscope optics.

To train the algorithm, 50 objects were randomly selected from each image yielding a total training dataset of 1,100 objects. These 1,100 objects were given two labels by the programmer to identify their group to the program. The first label indicated the color of the objects: Gram-positive, Gram-negative, “yeast-specific color”, or undetermined. The second label indicated the shape: “yeast-specific shape”, rod, cocci, or undetermined. The sub-images were labeled individually, in contrast to the slide label, because objects between groups often appear quite similar. For instance a cylindrical rod viewed in a perpendicular cross-section is a circle. Many of these smaller

sub-images received the label of undetermined because the class could not be determined.

After the training phase, the remaining 5,996 objects were used to test the two algorithms. Of these classified objects 258 had a label of undetermined for both the shape and color. Another 230 (4%) had inconsistent labels. For example, if the color was labeled as yeast and the shape did not indicate yeast then the two labels were considered inconsistent. Those objects with undetermined or inconsistent labels were excluded from the analysis. The removal of these points from analysis was justified because some of the sub-images did not represent microorganisms. For example, some sub-images were created by dust on the optics. Other sub-images may have truly represented microorganisms, but their true classification is difficult, likely impossible, to determine.

The true classification for the labeled objects was assumed to be the overall label for the slide from where the image was taken. For example if an image of Gram-positive cocci had 52 objects then all of those objects were assumed to be Gram+ cocci. An alternative approach would be to manually classify all of the test objects and compare the manual labels to the computer-generated labels, but this was not done for this study.

The sensitivity and specificity of each of the five traits are listed in Table 2 of the Tables section. The correct classification of a sub-image as Gram-positive as well as the classification of a sub-image as Gram-negative achieved a sensitivity of 99% and a

specificity of 99%. The classification of yeast reached 90% sensitivity and 99% specificity. The classification of rod had a sensitivity of 81% and a specificity of 78%. Finally, the identification of cocci had a sensitivity of 66% and a specificity of 87% for the set of sub-images.

Experiment 2:

In a second experiment, the program used in the first experiment was tested for the ability to classify a different larger and more diverse image library. This database contained 780 images taken from 78 slides. The database consisted of 15 Gram-negative rods, 46 Gram-positive cocci, 9 Gram-positive rods, and 8 yeast. No training data from the larger database was used to inform the program. Therefore, the only training data used by this program was the training data used to train the program in experiment one, namely 1,100 objects taken from 2 yeast, 4 Gram-positive cocci, zero Gram-negative cocci, 1 Gram-positive rod, and 4 Gram-negative rods. The outcome measure was the slide label. The slide label was determined by classifying the sub-images for 10 images per slide, and setting the slide class equal to the most common class of the classified objects.

The results show the program classified 60 out of 78 (77%) of the slides correctly. The confusion matrix for the classification is shown in Table 3. A confusion matrix tabulates the result predicted by the classifier to the true value.

Experiment 3:

The third experiment investigated the ability of the program to distinguish between classes given additional training. Using the large database of 780 images, a set of 3,000 sub-images were randomly chosen. A total of 750 objects were used from each of four classification groups: yeast, Gram-positive cocci, Gram-positive rods, and Gram-negative rods. The choice of random objects occurred anew for each of the five separate trials. The label assigned to the sub-image was the label of the slide from where the sub-image originated.

The training algorithm was configured to avoid memorizing the training data. The output of the classifier compared with the label of the original image determined if the classifier had correctly labeled the sub-image.

The results of the experiment are shown as a confusion matrix. The accuracy for the five trials ranged from 93% to 95% correct classification of the sub-image. The average across the five trials was 94% with the standard error of the mean equal to 0.21. The confusion matrices for the experiment are included in Table 4.

Discussion

The purpose of this thesis was to investigate the feasibility of microorganism classification using digital image analysis. To pursue this question a prototype program was created and tested against a library of images. The hypothesis stated that a software program could accurately classify microorganisms into major classes. The results of the first experiment demonstrated the ability to distinguish between major groups by classifying 5,500 unknown objects with high sensitivity and specificity. A second experiment applied the algorithm trained in the first experiment to a larger database. Reasons for its short-comings in the new setting include the limited quantity and diversity of its original training data compared to the dataset it attempted to classify. Another reason, disproved in the third experiment, could be an inherent inability of the learning algorithm to distinguish between the groups of the larger dataset. The third experiment demonstrated the learning algorithm could distinguish between the groups given sufficient training by classifying about 95% of the sub-images correctly. Thus the limitations of the second experiment are most likely the quality and/or quantity of the training data. Overall this project suggests that digital image analysis can classify microorganisms, but it will require training on large, diverse datasets that accurately represent the morphologic diversity in clinically relevant microorganisms.

The prototype program used in this feasibility study had a number of limitations affecting its overall performance. To begin it was trained using a set of only 22 images

representing 11 microorganisms. These microorganisms had low overall diversity because they included two yeast, four Gram-negative rods, one Gram-positive rod, four Gram-positive cocci, and no Gram-negative cocci. In practical terms this smaller set of images did not fully captured the differences in morphology of the bacteria and yeast in the testing data set. When examining misclassified slides from experiment 2, the microorganisms had characteristics much different from those used in the training dataset. As an example, the program misclassified two examples of *Micrococcus*, an organism not present in the training dataset and morphologically quite different from other Gram-positive cocci. Because the training dataset was deficient in many microorganisms seen in the test, its ability to accurately classify the larger test dataset was limited.

The differentiation of Gram-positive cocci into subgroups by morphology is routine in the interpretation of a Gram stain. The morphologic distinction of “clusters” vs. “chains” is commonly thought to separate pathogenic Staphylococcus and Streptococcus/Enterococcus respectively. When these features were detected by the program they did not correlation with the final identification of subgroups of cocci. The interpretation of a Gram-smear as containing “clusters” or “chains” is useful because it implies a distinction between cocci. Because the detection of these features did not provide any value for sub-classification of Gram-positive cocci in this instance, it was excluded from the study.

In clinical practice, microtechnologists use the catalase test and morphology to distinguish between Gram-positive cocci when interpreting Gram-stained slides from cell culture plates. Unlike the typically longer chains found in blood culture smears, chain length from cell culture plates is variable. In addition, multiple viewings of a particular slide can disrupt clusters to form chains, which was the case with the slides viewed in this experiment. After controlling for these factors a future study should address this important area.

An improvement to the protoype algorithm could occur by incorporating the current test data into its training dataset. Different methods for how to incorporate the new data depend on the strategy chosen by the programmer. In situations with limited data and computing power, the best choice would be to curate the dataset to remove controversial points. Controversial points exist at the boundaries between groups. If these points are labeled, it can create extra work for the classifier to figure out where to place the decision boundary. In situations with limited computing power this can add significant time to the program. This type of approach was used to prototype the program because of limited data and computing power. A major downside to the use of curated data is the time required by the user to classify the data. Another downside to curated data would be the constant requirement for human intervention to retrain the program on new data. This would be an issue for a very large database.

An alternative strategy to incorporate the new data would be to label all the microoganisms from a slide as belonging to the category of the whole slide. The

drawback of this method is the increased computing time required to place the boundary. For example, every image of a rod contains objects that resemble cocci because looking down the long axis of cylindrical rod is circular, and because many rods become short at the time of cell division. The placement of a decision line between these groups with overlapping points was computationally prohibitive on my commodity laptop. Fortunately relatively inexpensive access to powerful computers is available. For instance Amazon and Google both offer cloud computing options capable of running this type of program.

Another improvement to the algorithm would be the incorporation of image level characteristics. In the current algorithm parameters are generated solely from the microorganisms themselves and not from the characteristics of the image. In essence each microorganism undergoes classification independent of near-by organisms. This choice was consciously made to avoid the assumption that the "representative" images taken from the slides actually contained information to distinguish between classes. For example images of yeast could be taken at low organism density compared to images of Gram-negative rods that were only taken at high organism density. Image-level information can offer significant advantages over classification parameters calculated solely from the individual organisms. For example, isolated cocci can provide ideal size and color characteristics that improve the algorithm's ability to identify less than ideal cocci, which may be embedded in sputum or grouped together in clumps. In a practical example of the use of image-level information to classify two groups, the receiver-

operator curve for a single size parameter increased considerably when information from surrounding organisms was included in its calculation. This is shown in Figure 2. As a second example, the crystal violet stain may not be equally retained by all the microorganisms in the slide. Therefore the color classifier for Gram-positive may require an adjustment based on the more distinctly Gram-positive organisms in the slide.

In the pursuit of automating the capture of images, a computer-controlled movement system was designed to fit on the current microscope. The system uses a microcontroller to control the movement of three small servo motors. The motors are attached to the three planes of the microscope to move the slide and focus the image. The microcontroller was successfully linked with the image analysis program to provide communication between the image analysis and movement of the slide. The model drawing for the system is shown in Figure 3. The system was built for well under $100 in raw materials. In the future a dedicated system for slide interpretation would be needed to scan the slide for interpretable images.

The majority of the project focused on classification of objects because the numbers of available slides was small. As a future project, slide classification using object and image level parameters would make the system more practical. A simple approach to slide classification would give each object a single vote towards a class with the class receiving the most points given as the slide name. As a more complex and realistic interpretation, objects could receive a weighted score depending on the

distance of their parameters from the decision boundary. Objects sitting closer to the decision boundary would be considered controversial and receive a lower score compared to objects farther away from the boundary. This could prevent slide misclassification in a scenario where a large number of controversial objects outvotes a smaller number of easily classified objects.

In the future, a confidence estimate for slide classification would be necessary to signal slides for operator review. This type of analysis would require a much larger database of slides than currently available. The morphology of microorganisms has numerous modifying variables before the sample reaches the laboratory as well as after it is processed. Variables affecting morphology before the sample reaches the laboratory include site of infection, patient antibiotic use, transport time, transport conditions, and potentially regional variation in microorganisms. Once the sample has reached the lab, culture duration, culture media, and incubator conditions could affect morphology. Calculation of the relative importance of these variables necessitates a large and diverse slide database. Thus the creation of a slide database is fundamentally important to the operation of a digital image analysis program.

Many questions remain open in the application of digital image analysis to microbiology. For instance, how much taxonomic information can morphology provide? Do differences in slide preparation affect digital image interpretation? How extensible is single software approach to the variety of microscope tasks? How biased is a data curator's opinion of a microorganism's label? Will image segmentation work in a more

complex background? How much will a system cost and how long will it take to give a result? The field is essentially wide open.

In conclusion, digital image analysis is an FDA approved technique for the interpretation of medical images in other areas that has potential applications in clinical microbiology. Overall, the prototype algorithm trained on a modest learning dataset was able to classify most slides in a larger slide database, which strongly suggests digital image analysis is a feasible approach to microbiology automation. The foundation for future work requires a system capable of automated image capture and a large database of slides representing the diverse variables affecting microorganism morphology. Many important questions in the application of digital image analysis to microbiology remain unanswered making this area fertile for future research.

References

1. Craig, A.G., and Hoheisel, J.D., editors. 1999. *Automation: Genomic and Functional Analyses*: Academic Press. 4-6 pp.
2. Murray, P.R., Jorgensen, E.J.B.J.H., Landry, M.L., and Pfaller, M.A., editors. 2007. *Manual of Clinical Microbiology*: ASM Press. 182-186 pp.
3. 2001. Evaluation of VITEK 2 rapid identification and susceptibility testing system against gram-negative clinical isolates. *J Clin Microbiol.* 39:2964-2966.
4. Ward-Cook, K., and Tannar, S. 2001. 2000 Wage and Vacancy Survey of Medical Laboratories. *Laboratory Medicine* 32:124-138.
5. Steward, C.A., Ward-Cook, K., and Tannar, S. 2005. 2003 Wage and Vacancy Survey of Medical Laboratories. *LabMedicine* 36:149-157.
6. Bennett, A., and Thompson, N.N. 2009. ASCP Wage and Vacancy Survey of U.S. Medical Laboratories. *LabMedicine* 40:133-141.
7. Pathology, A.S.f.C. 2007. ePolicy News. American Society for Clinical Pathology.
8. Administration, U.S.F.a.D. 2011. 510 (k) Premarket Notification.
9. 2009. Multi-modality CADx: ROC study of the effect on radiologists' accuracy in characterizing breast masses on mammograms and 3D ultrasound images. *Acad Radiol.* 16:810-818. Epub 2009 Apr 2017.
10. 1999. Determining the utility and effectiveness of the NeoPath AutoPap 300 QC System used routinely. *Acta Cytol.* 43:13-22.
11. May, M. 2010. A Better Lens on Disease. *Scientific America* 302:74 - 77.
12. Baron, E.J., D'Souza, H., Wang, A.Q., and Gibbs, D.L. 2008. Evaluation of the Biomic V3 Microbiology System for Identification of Selected Species on BBL CHROMagar Orientation Agar and CHROMagar MRSA Medium. *American Society for Microbiology* 46:3488-3490.
13. 2010. Grid infrastructures for developing mammography CAD systems. *Conf Proc IEEE Eng Med Biol Soc* 2010:3467-3470.
14. Craig, J.J. 2004. *Introduction to Robotics: Mechanics and Control*: Prentice Hall.
15. 2010. Mass spectrometry tools for the classification and identification of bacteria. *Nat Rev Microbiol.* 8:74-82.
16. Baxt, W.G., and Skora, J. 1996. Prospective validation of artificial neural network trained to identify acute myocardial infarction. *The Lancet* 347:12-15.
17. Adam, B.L., Qu, Y., Davis, J.W., Ward, M.D., Clements, M.A., Cazares, L.H., Semmes, O.J., Schellhammer, P.F., Yasui, Y., and Feng, Z. 2002. Serum protein fingerprinting coupled with a pattern-matching algorithm distinguishes prostate cancer from benign prostate hyperplasia and healthy men. *Cancer Research* 62:3609.
18. Hong, P., Liu, X.S., Zhou, Q., Lu, X., Liu, J.S., and Wong, W.H. 2005. A boosting approach for motif modeling using ChIP-chip data. *Bioinformatics* 21:2636.
19. Li, J., Zhang, Z., Rosenzweig, J., Wang, Y.Y., and Chan, D.W. 2002. Proteomics and bioinformatics approaches for identification of serum biomarkers to detect breast cancer. *Clinical Chemistry* 48:1296.
20. Huang, K., and Murphy, R.F. 2004. Boosting accuracy of automated classification of fluorescence microscope images for location proteomics. *BMC bioinformatics* 5:78.
21. Kim, K.J., and Cho, S.B. 2003. Evolving artificial neural networks for DNA microarray analysis. IEEE. 2370-2377 Vol. 2374.

22. Lin, C.C., Tsai, Y.S., Lin, Y.S., Chiu, T.Y., Hsiung, C.C., Lee, M.I., Simpson, J.C., and Hsu, C.N. 2007. Boosting multiclass learning with repeating codes and weak detectors for protein subcellular localization. *Bioinformatics* 23:3374.
23. Valetudie, G., Desachy, J., and Sola, C. 2007. Markov Models to classify M. tuberculosis spoligotypes. IEEE. 668-671.

Figures

Figure 1: The process of segmentation. (Left) Original image (Right) Segmented image

Note: The images are displayed at different scales.

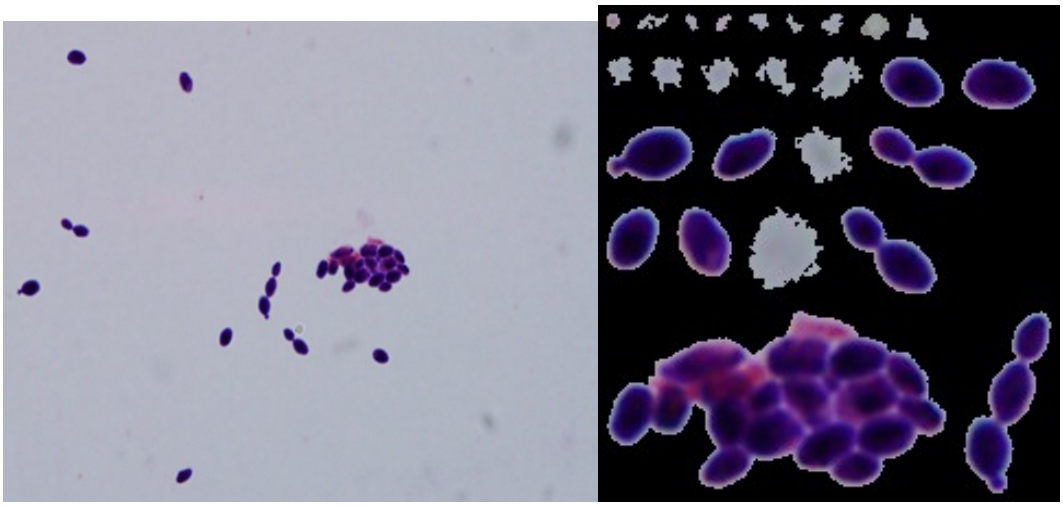


Figure 2: An image level parameter uses information from the entire image to calculate a parameter for an object, while an object level parameter calculates the parameter independent of other objects within the image. This image level parameter (blue curve) outperforms the object level parameter (red curve).

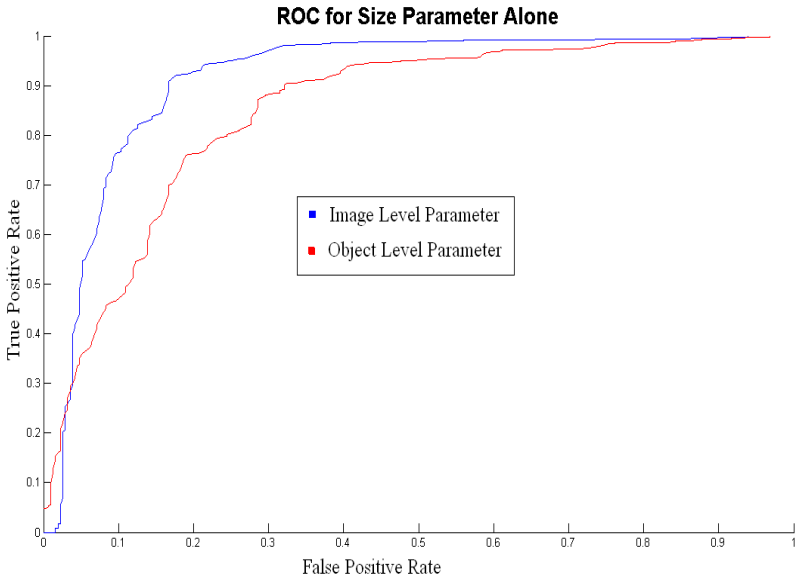
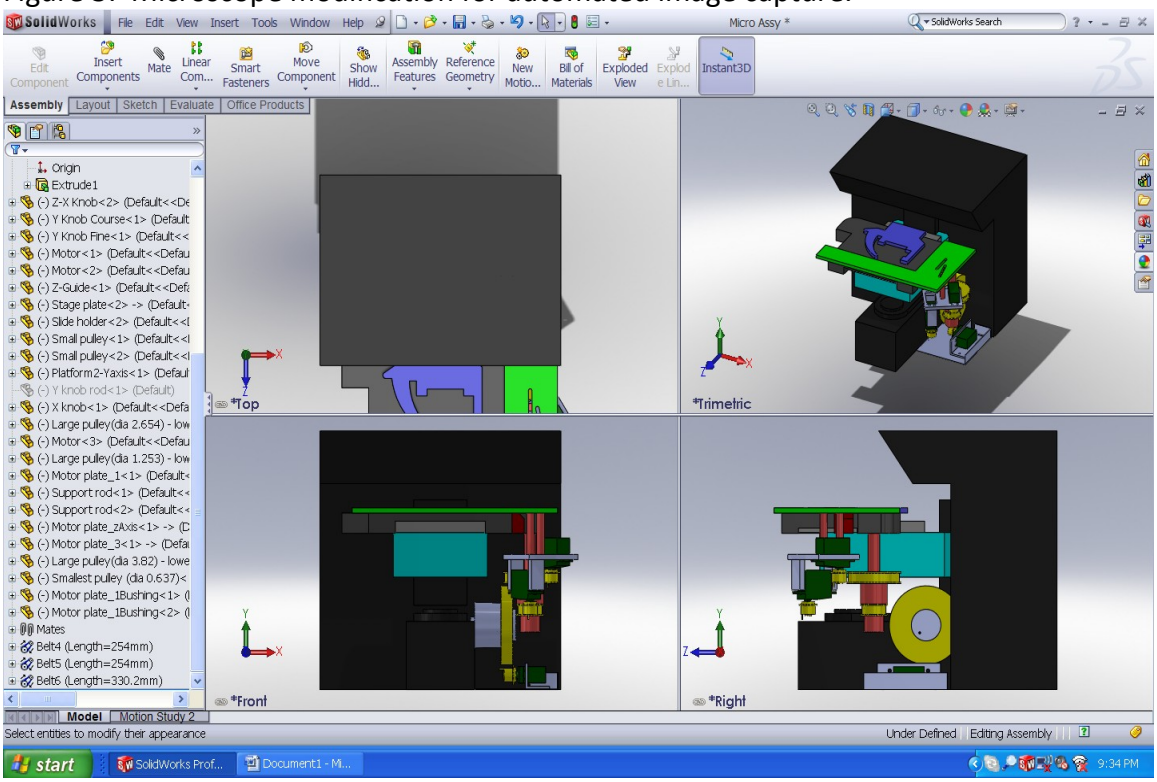


Figure 3: Microscope modification for automated image capture.



Tables

Table 1: Slide Database Catalog (GPC: Gram-positive cocci; GPR: Gram-positive rod, GNR: Gram-negative rod; NA: Not available)

Catalog Number	Gram Stain Interpretation	Species
1	GNR	Serratia marcescens
2	GPC	Enterococus faecalis
3	GNR	Klebsiella pnemonae
4	Yeast	Candida glabrata
5	GNR	Enterobacter cloacae
6	GPC	Streptococcus bovis
7	GNR	Escherichia coli
8	GNR	Pseudomonas aeruginosa
9	GPC	Micrococcus
10	GPC	Enterococcus faecium
11	GNR	Psuedomonas aeruginosa
12	GPC	Enterococcus faecalis (VRE)
13	Yeast	Candida glabrata
14	GPC	Enterococcus faecalis
15	GPC	Entercoccus faecalis
16	GPC	Micrococcus
17	GNR	Pseudomonas aeruginosa
18	GPC	Staphylococcus - Coagulase negative
19	Yeast	Candida parapsilosis
20	GNR	Escherichia coli
21	GNR	Proteus mirabilis
22	GNR	Haemophilus influenzae
23	GNR	Citrobacter freundii
24	Yeast	Candida albicans
25	GPC	Staphylococus - Coagulase negative
26	GPC	NA
27	GPC	Staphyloccoccus lugdunensis
28	GPC	Staphylococcus epidermidis
29	GPC	Staphylococcus aureus
30	GPC	Streptococcus viridans
31	GPC	Streptococcus viridans
32	GPR	Lactobacillus
33	GPR	Corynebacterium striatum
34	GPC	Staphylococcus - Coagulase negative
35	GPC	Streptococcus agalactiae

36	GPR	NA
37	GPC	Staphylococcus - Coagulase negative
38	GPC	Staphylococcus - Coagulase negative
39	GPC	NA
40	GPC	NA
41	GPR	NA
42	GPC	Staphylococcus aureus
43	GPC	Staphylococcus aureus
44	Yeast	NA
45	Yeast	NA
46	Yeast	NA
47	GPR	Propionibacterium
48	GPC	NA
49	GPC	Enterococcus
50	GPC	Staphylococcus aureus
51	GPC	NA
52	GPC	Staphylococcus - Coagulase negative
53	GPC	Staphylococcus aureus
54	GPC	NA
55	GPC	NA
56	GPC	NA
57	GPC	NA
58	GPR	NA
59	GPC	NA
60	GNR	Haemophilus influenzae
61	GNR	NA
62	GNR	Prevotella oralis
63	GPR	Staphylococcus - Coagulase negative
64	GPC	Staphylococcus aureus
65	GPC	Staphylococcus - Coagulase negative
66	GPC	Staphylococcus aureus
67	GPR	NA
68	GPC	Staphylococcus aureus
69	GPC	Staphylococcus aureus
70	GPR	Propionibacterium
71	GPC	NA
72	GNR	NA
73	GPC	NA
74	GPC	NA
75	GPC	NA
76	GPC	NA
77	GPC	NA

78	Yeast	NA

Table 2: Experiment 1 Sensitivity and Specificity of Binary Classifiers for Microorganism Traits (N = 5508). (488 points removed for undetermined or inconsistent labels.)

	Sensitivity	Specificity
Yeast	1.00	0.98
Cocci	0.94	0.84
Rods	0.89	0.93
Gram-Positive	0.99	0.99
Gram-Negative	0.99	0.99

Table 3: Experiment 2 Confusion Matrix For Slide Classification

(GPC: Gram-positive cocci; GPR: Gram-positive rod, GNR: Gram-negative rod)

		Predicted				
		Yeast	GPC	GNC	GPR	GNR
Known	Yeast	4	0	0	0	4
	GPC	0	35	2	1	8
	GNC	0	0	0	0	0
	GPR	0	0	0	6	3
	GNR	0	0	0	0	15

Table 4: Experiment 3 Confusion Matrices for 3000 Objects Classified Over 5 Trials

(GPC: Gram-positive cocci; GPR: Gram-positive rod, GNR: Gram-negative rod)

Trial #1

		Predicted			
		Yeast	GPC	GPR	GNR
	Yeast	690	25	23	12
	GPC	19	684	31	16
Known	GPR	10	23	712	5
	GNR	9	16	10	715

Trial #2

		Predicted			
		Yeast	GPC	GPR	GNR
	Yeast	695	21	19	15
	GPC	14	698	22	16
Known	GPR	17	15	714	4
	GNR	9	7	5	729

Trial #3

		Predicted			
		Yeast	GPC	GPR	GNR
	Yeast	700	28	9	13
	GPC	19	689	24	18
Known	GPR	9	26	706	9
	GNR	10	10	10	720

Trial #4

		Predicted			
		Yeast	GPC	GPR	GNR
	Yeast	712	19	11	8
	GPC	19	677	33	21
Known	GPR	15	9	721	5
	GNR	12	11	8	719

Trial #5

		Predicted			
		Yeast	GPC	GPR	GNR
	Yeast	701	21	19	9
	GPC	20	668	42	20
Known	GPR	9	19	717	5

GNR	10	8	7	725