B657 Computer Vision
Final Project: Pharmaceutical Pill Recognition Using Computer Vision Techniques
Spring 2016

Charlene Tay (ctay)
Mridul Birla (mbirla)

May 5, 2016

1. Introduction

The National Library of Medicine (NLM) recently put out a "Pill Image Recognition Challenge"[1], seeking algorithms and software to match images of prescription oral solid-dose pharmaceutical medications (pills, including capsules and tablets). For our final project in B657 Computer Vision, we have decided to take up this challenge. Such algorithms and software can be applied in many important ways, especially in helping medical care professionals and their patients deal with unidentified or mislabeled prescription pills. Senior citizens are especially affected by this, with nine out of 10 of US citizens over the age of 65 who take more than one prescription pill being prone to misidentifying their pills. Taking such pills can result in adverse drug reactions that affect health or could even cause death. By coming up with ways to easily identify and verify prescription pills, errors can be greatly reduced. Another useful application of pill recognition would be in aiding law enforcement with the identification of counterfeit or illicit drug pills.

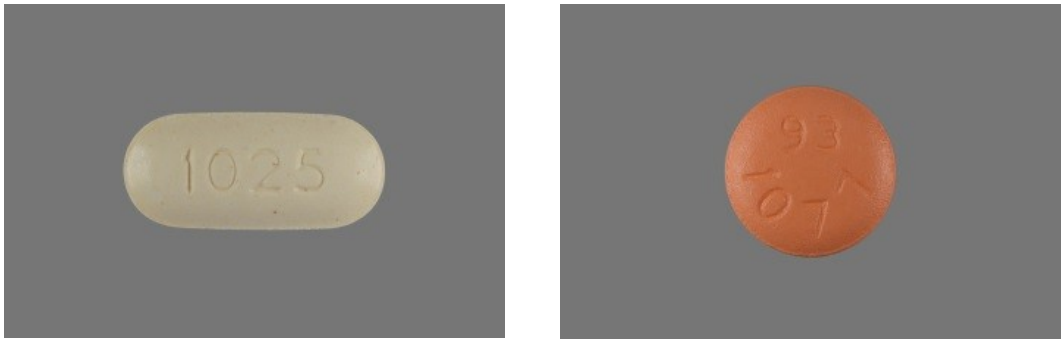

Figure 1: Example reference images of pills provided in the dataset

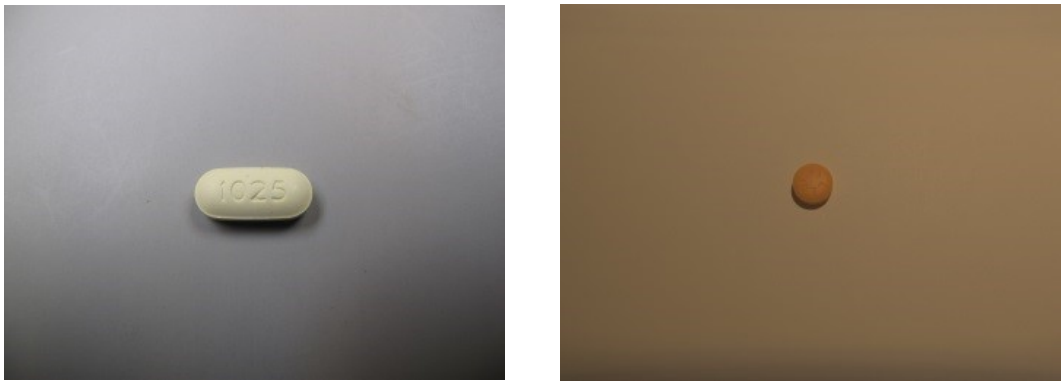

Figure 2: Consumer quality images of the same two pills shown in Figure 1

[1] http://pir.nlm.nih.gov/challenge/

2. Overview

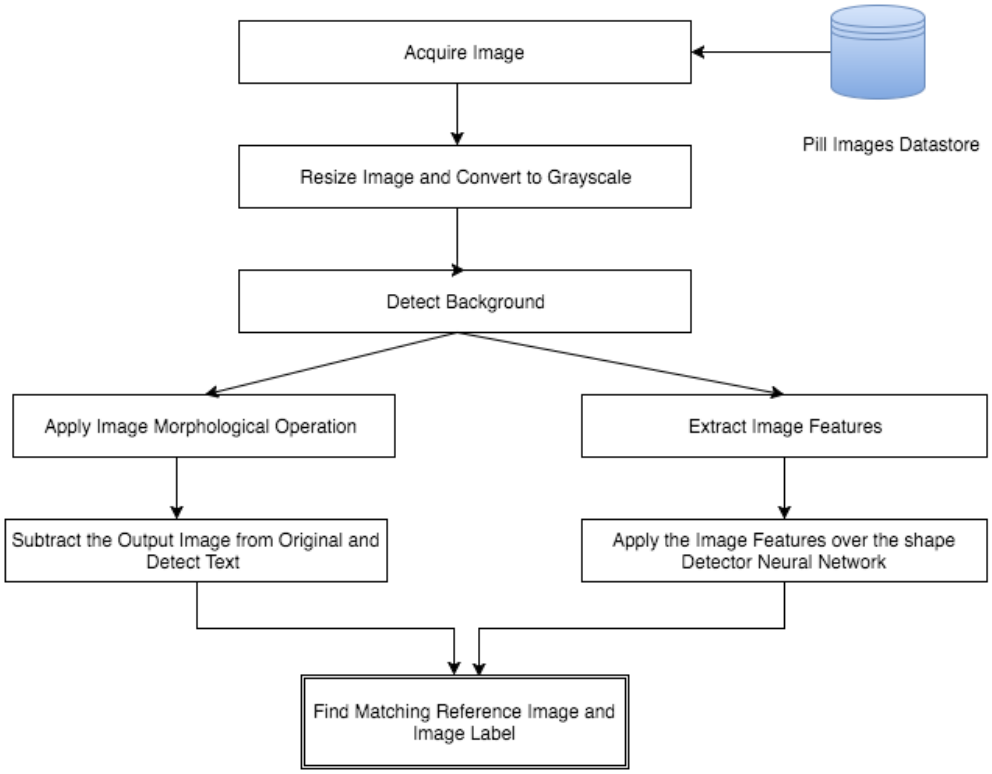

Figure 3: Overview of Steps Taken

For use in designing the submission to this Challenge, the National Library of Medicine (NLM) has provided 7000 images of pills taken as part of the Computational Photography for Pill Identification Project[2]. This set of images consists of 2000 JPEG reference images (one for the front and one for the back of each of 1000 pills), 5000 JPEG consumer-quality images of the same 1000 pills, and a ground truth table. The reference images are high resolution photographs of 1000 prescription pills taken under laboratory lighting conditions, from a camera directly above the front and the back faces of the pill, and using specialized digital macro-photography techniques. The consumer-quality images, on the other hand, are photographs of the same 1000 pills but taken with digital cameras such as the ones in mobile devices, and taken under different conditions. Please refer to Figure 1 and 2 for examples of these images.

By using the image processing methods learnt over the course, the aim of this project was to implement a system that would take the consumer images of their prescription pills and identify them (or suggest most likely matches) by comparing them against the image database of pills provided by the NLM and matching them to the reference image

[2]https://lhncbc.nlm.nih.gov/project/c3pi-computational-photography-project-pill-identification

of the same pill. In particular, we were interested in using techniques for extracting features such as pill shape and imprinted identifying numbers.

Initial ideas for this project included using color as a factor to determine pill identity. However, after working with the test images, it became evident that using color would not be a good way to go. Variation in the lighting conditions and the use of different camera lenses for the capturing of the consumer-quality images results in a large difference and variation in the detected RGB values of each pill. We thus decided to focus our efforts on shape detection and imprint extraction.

Figure 3 is a flowchart depicting the steps that have been taken towards processing the images for pill identification. Upon acquiring the images, each image is resized and converted to grayscale. Following that, image segmentation is performed to separate the pill from the background. From here, the image goes through two waves of processes - one for pill imprint/feature extraction and detection, and one for shape detection. This report will describe the steps that we have taken towards this goal and describe the results that we have obtained from trying to evaluate the performance of these methods.

3. Preprocessing

The first few weeks of working on this project was spent mostly on trying to finalize the process and figuring out methods that would work for this project. We read through the papers that were previously listed in the project proposal for ideas on methods that we could implement towards this goal. Our focus so far has been on extracting features for determining pill shape. We have worked on preparing a neural network for classifying the images according to shape. A good portion of our time has been spent on constructing code for implementing the neural network.

Next, we had to prepare our training data for the neural network. The ground truth table provided by the NLM maps each of the consumer quality pill images to the name of its corresponding reference image. While this is great for evaluation purposes, the ground truth table does not give us any other information about the pills. In order to train our neural network to classify each pill image according to shape, we had to first code and label each training image according to their pill shape. There are 1000 pills in the image set, so it took us some time to manually classify and label each pill shape.

In terms of image pre-processing for pill shape detection, the methods that we initially tried were not very successful. We did not have desirable results using the Canny edge detector [1], and applying the Sobel operator[3] for edge detection was not perfect either. Other methods that we tried were to implement histogram equalization. For this, we tried both global histogram visualization and Contrast Limited Adaptive Histogram Equalization using the OpenCV library[4]. Example results of this process can be seen in Figure 4. These methods, however, did not prove to be useful for our purposes.

Our current flow of processing includes converting the images to grayscale, conducting background subtraction, followed by applying affine transformation to rescale the

[3]http://homepages.inf.ed.ac.uk/rbf/HIPR2/sobel.htm
[4]http://opencv-srf.blogspot.com/2013/08/histogram-equalization.html

Figure 4: Image after Adaptive Histogram Contrast and Global Histogram Equalization

images. To pre-process the images for imprint detection and extraction, we tried various methods and had the most success with applying morphological image processing operations on the images to extract the imprints. The original image is put through an erosion technique [5]. The erosion of the image A by the structuring element B is defined by:

$$A \ominus B = z \in E | B_z \subseteq A$$

where E is a Euclidean space or an integer grid, A an image in E, and B_z is the translation of B by the vector z. We use erosion for edge detection by taking the erosion of an image and then subtracting it away from the original image, thus highlighting just those pixels at the edges of objects that were removed by the erosion. The edge detection process also allows us to create an image mask of the outline of each pill, allowing us to calculate shape features to aid in shape detection.

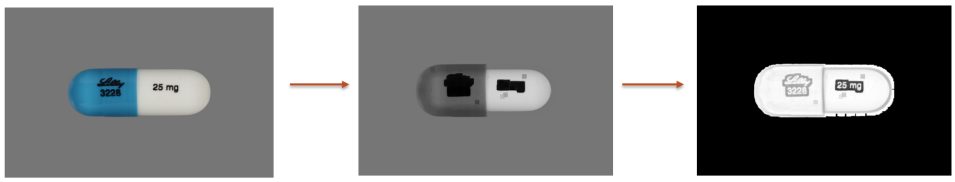

Figure 5: Left: Original Image; Middle: Image after applying erosion techniques; Right: Output after subtraction erosion output image from original image

4. Imprint Detection and Extraction

Pill shape, size and color are limited in variation, so the most distinguishing features that will help in identifying each pill are the imprints or markings on the pills themselves. To aid us in this, we researched some optical character recognition (OCR) implementations to automatically detect the characters imprinted on the pill. There are many free optical OCR software available online for use. However, we had very little success with most of these free software. We did however get good results from one website[6] on capsules that

[5] http://homepages.inf.ed.ac.uk/rbf/HIPR2/erode.htm
[6] https://ocr.space/

had markings imprinted in black text on the capsules. We had no success at all with using the OCR software on pills with engraved imprints.

Figure 6: Example input and output from online OCR

We decided to move forward and focus more on the shape detection, but in the near future we hope to be able to find and implement a better way for extracting the imprinted or engraved features on each pill.

5. Shape Detection

One of our main interests for the final project was to be able to get some experience with the implementation and usage of neural networks. Classification of pills by fine-tuning a deep convolutional neural network was quickly swept under the rug as an option for this problem, as the identification of individual pill types is too specific for that application. A fine-tuned deep convolutional neural network would be able to distinguish pills from other categories of objects, but it would not work well for distinguishing specific pills amongst other pills. As such, in order for us to still get experience with neural networks, we decided to try implementing our own back-propagation neural network for shape classification.

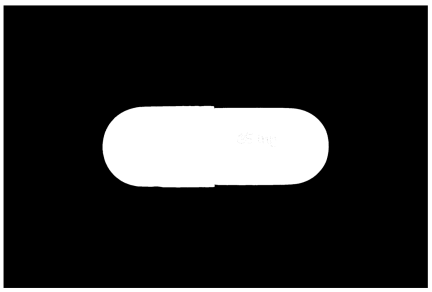

Figure 7: Example of image mask used to calculate shape features

Our pre-processing steps result in an image mask for each pill. This gives us binary information that separates the background from the pill. Using this information, we can calculate measures such as perimeter of the shape, area, and other measures. As suggested in a paper that we have read regarding shape classification using neural

networks [6], we calculated various measures from each blob to that would help to determine shape. To determine pill shape, twelve features of image shape are extracted from each blob. These include seven Hu invariant moments, circularity degree, rectangularity degree, sphericity degree, concavity degree and flat degree. We were able to obtain these measures by using various operations in the open-source libraries, OpenCV[7] and SimpleCV[8].

5.1 Hu Moments

One measure that we decided to use was "image moments", which are two-dimensional statistical moments. To calculate these moments, the following equation is used:

$$\mu_{i,j} = \frac{\Sigma_{x=1}^{W}\Sigma_{y=1}^{H}(x-\mu_x)^i(y-\mu_y)^j f(x,y)}{\Sigma_{x-1}^{W}\Sigma_{y-1}^{H} f(x,y)}$$

where f(x,y) is the pixel value at coordinates (x,y), and W and H are the image width and height. μ_x and μ_y indicate the average x and y. The values i and j select the order of moment that we want to compute, and various orders of moments can be combined in different ways. In particular, we decided to compute the seven Hu moments [4] using combinations of image moments:

$$\begin{aligned}
H_1 &= \mu_{20} + \mu_{02} \\
H_2 &= (\mu_{20} - \mu_{02})^2 + 4\mu_{11}^2 \\
H_3 &= (\mu_{30} - 3\mu_{12})^2 + 3(\mu_{21} - \mu_{03})^2 \\
H_4 &= (\mu_{30} + \mu_{12})^2 + (\mu_{21} + \mu_{03})^2 \\
H_5 &= (\mu_{30} - 3\mu_{12})(\mu_{30} + \mu_{12})[(\mu_{30} + \mu_{12})^2 - (\mu_{21} - \mu_{03})^2] \\
H_6 &= (\mu_{20} - \mu_{02})[(\mu_{30} + \mu_{12})^2 - (\mu_{21} + \mu_{03})^2 + 4\mu_{11}(\mu_{30} + \mu_{12})(\mu_{21} + \mu_{03})] \\
H_7 &= 3(\mu_{21} - \mu_{03})(\mu_{30} + \mu_{12})[(\mu_{30} + \mu_{12})^2 - 3(\mu_{21} + \mu_{03})^2] \\
&\quad +(\mu_{03} - 3\mu_{12})(\mu_{21} + \mu_{03})[3(\mu_{30} + \mu_{12})^2 - (\mu_{21} + \mu_{03})^2]
\end{aligned}$$

Hu moments are useful for our purposes because they are scale and rotation invariant, making them robust affine invariant image descriptors.

5.2 Other Shape Descriptors

As suggested in the paper regarding shape classification using neural networks [6], here are the other shape measures that we chose to extract and use as shape descriptors:

- Circularity degree is also referred to as shape factor, a function of the perimeter P and the area A:

$$f_{circ} = \frac{4\pi A}{P^2}$$

[7] http://opencv.org/
[8] http://simplecv.org/

To obtain this, we calculated the ratio of the square of the perimeter of the shape, to the area of the shape.

- Rectangularity degree, also known as the rectangle fitting factor, is defined as the ratio of the area of the shape and the area of minimum enclosing rectangle of the shape.
- Sphericity degree is defined as the ratio of the radius of internally tangent circle of the shape, and the radius of circumcircle of the shape.
- Concavity degree is defined by the following equation:

$$y = 1 - \frac{s_0}{s_r}$$

where s_0 is the area of approximation polygon of the shape, and s_r is the area of the minimum external convex.

- Flat degree can be calculated by finding the ratio of the length of the shape's long axis by the length of its short axis. For example, the flat degree of a rectangle is its aspect ratio - the ratio of the rectangle's length by it's width.

5.3 Neural Network

The above measures were calculated for each blob/shape. These measures were then sent as an input vector to a neural network that we implemented.

Figure 8: Examples for each of our shape categories

The neural network created was 3 layered with 200 hidden nodes and 9 output nodes which was trained with 10 iterations. For training, we first manually categorized and labeled according to shape each of the images provided by the NLM. We first

categorized these images into 20 distinct shapes. However, as a few of these shapes were extremely underrepresented in our dataset, we decided to group some of these together for better training. For example, categories that were very close to each other in shape like football, oval and oblong were clustered together for training in the neural network. After regrouping the categories, we were left with nine shape categories for shape classification by the neural network.

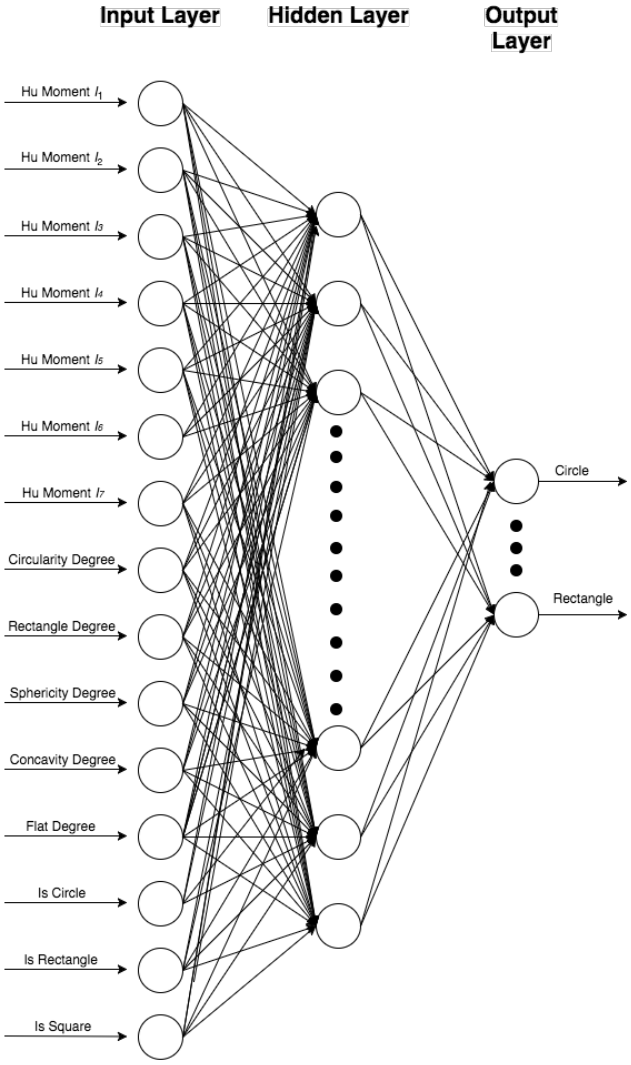

Figure 9: Diagram of our Neural Network

6. Results

Once the neural network was trained, we fed in the test images into the neural network for evaluation. The overall accuracy of our neural network for classifying pill shapes was 50.7%. We were surprised by this and tried to take a closer look at the results by looking at each of the shape categories. Figure 10 is a graph showing the accuracy of the neural network for classifying each of our shape categories. As seen in the graph, we had good accuracy for shapes such as the circle (75%), and oblong/football/oval (70%), but very low accuracy for other shapes like triangle (10%) and square (15%).

This is a surprising result because these shapes are very distinct to humans (if it has three sides, it should be a triangle). One reason why we think that we achieved poor results for this is because the number of instances that these shapes appear in our data set is very low. For example, the triangle pills only appear once or twice in our set of images.

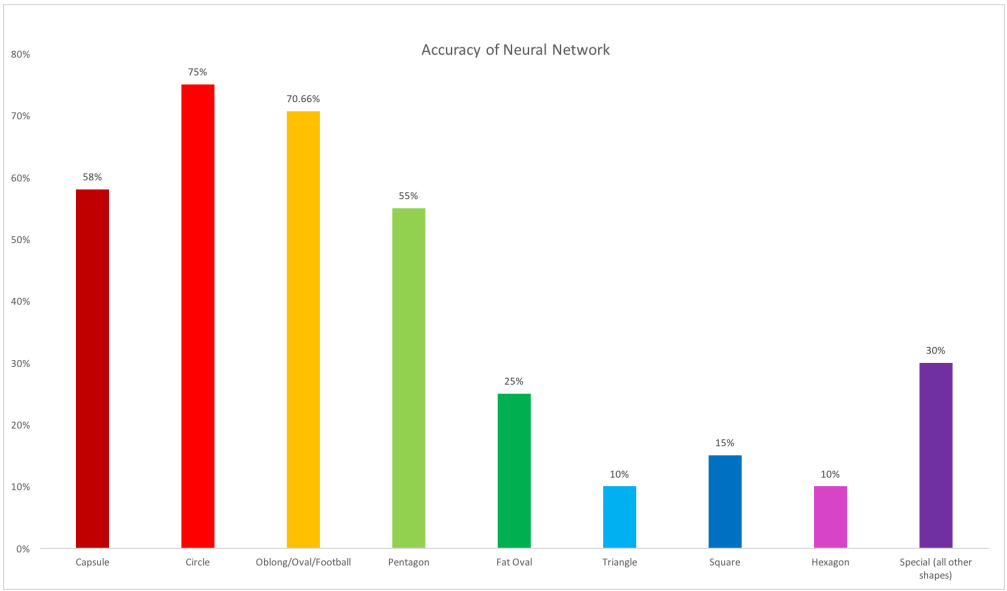

Figure 10: Graph showing the accuracy of the Neural Network for classifying each shape category

For further evaluation, we decided to try using a number of classifiers to see if we can get a good accuracy for shape image classification using the shape measures that we have extracted. We fed our images and the shape descriptors into classifiers such as k-Nearest Neighbor, Support Vector Machine, Decision Tree, Random Forest, and Adaboost. The labels used for input into these classifiers were the 20 shapes that the images were originally categorized into, not the nine shape groups that were used to train the neural network. Despite the addition of categories for classification, these other classifiers generally performed better than the neural network.

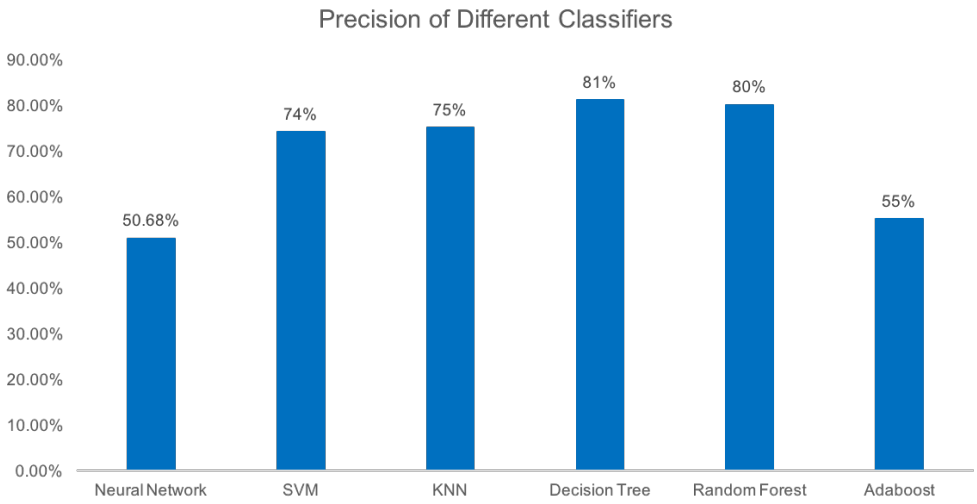

Figure 11: Graph showing the accuracy of each classifier for classifying each shape category

As shown in Figure 11, the other classifiers worked better overall than the neural network and classified the pill shapes with higher precision. The decision tree worked best as a shape classifier, with a precision of 81%. Random forest, k-Nearest Neighbor classifier, and the Support Vector Machine classifiers also performed well, with precisions ranging from 74-80% as compared to the 50% achieved from the neural network.

7. Discussion

The problem of pill recognition using Computer Vision has been a very interesting one to think about, and we have enjoyed working on this as a project. The task turned out to be a lot more difficult than we first envisioned, and we hit many roadblocks in regards to trying to identify the pills by features such as color and imprints.

The challenges in this task were mostly due to limitations in the amount and type of data. First, classification and identification of pill color was not a practical option for us due to differing light conditions and camera lenses used in the consumer-quality test images. These varying lighting conditions and the different scales at which the photographs were taken also made features such as pill engravings and imprints very difficult to extract. In addition, the dataset contained 1000 different pills to distinguish between, but only a few copies of each individual pill.

While we were successful with using shape descriptors and various classifiers to classify pills by shape, our initial idea of using a neural network did not work out as well as expected. This was also because of the limited amount of data, as well as a lack of diversity of shape in the data. Many of the shapes were very underrepresented in the dataset, and the neural network did not get trained properly on these shapes. Nonetheless, we gained experience with using neural networks and other classifiers. We are quite happy

with the precision that we are getting for classifying the pills by shape using the shape descriptors that we have calculated for each pill image. These classifiers, especially the decision tree (best performance), can run quickly and would be good for implementation in applications for real time pill recognition.

In the near future, we hope to be able to continue working on this problem and find better solutions to solve it so that we may work towards a viable entry in the NLM Pill Recognition challenge. Methods to be considered include using Multi-scale Local Binary Pattern (MLBP) and Scale Invariant Feature Transform (SIFT) to find the keypoints mapping pill imprinting, as suggested by Lee and coworkers [5]. It might also be helpful to train our system with images from different light exposures and camera distances, as described by Cunha and colleages [2]. Alternatively, Hartl et al [3] determined color by applying a white balance algorithm to deal with light variations, then determined color by comparing their color histogram with a pre-build lookup table based on Identa database. It was also suggested to us recently that we might try using Active Contours (Snakes)[9] to segment the image. There are many other methods out there, but this is still an open problem that researchers are actively trying to resolve. We are excited to keep working towards a working application for pharmaceutical pill recognition that would help the general population.

[9] http://www.iacl.ece.jhu.edu/static/gvf/

Bibliography

[1] John Canny. A computational approach to edge detection. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 8(6):679–698, 1986.

[2] A. Cunha, T. Adao, and P. Trigueiros. Helpmepills: a mobile pill recognition tool for elderly persons. *Procedia Technology*, 16:1523–1532, 2014.

[3] A Hartl, C Arth, and D Schmalsteig. Instant medical pill recognition on mobile phones. In *In Proceedings of the IASTED International Conference on Computer Vision*, pages 188–195, 2011.

[4] Ming-Kuei Hu. Visual pattern recognition by moment invariants. *IRE Trans. Info. Theory*, IT-8:179–187, 1962.

[5] YB Lee, U. Park, Anil Jain, and SW Lee. Pill-id: Matching and retrieval of drug pill imprint images. In *20th International Conference on Pattern Recognition (ICPR)*, pages 2632–2635, 2010.

[6] Wenpeng Lu. Method for image shape recognition with neural network. In *Advances in Computer Science and Information Engineering*, pages 547–551. Springer Berlin Heidelberg, 2012.