IMAGE PROCESSING BASED FABRIC DETECTION

BY

MD. JAOWAD HASAN
ID: 151-15-5495

This Report Presented in Partial Fulfillment of the Requirements for the Degree of Bachelor of Science in Computer Science and Engineering

Supervised By

ANUP MAJUMDER
Lecturer
Department of CSE
Daffodil International University

Co-Supervised By

MASUD RABBANI
Lecturer
Department of CSE
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

DECEMBER 2018

APPROVAL

This Research titled "**Image Processing Based Fabric Detection",** submitted by **Md. Jaowad Hasan,** ID No: **151-15-5495** to the Department of Computer Science and Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on 10/12/2018.

BOARD OF EXAMINERS

____

Dr. Syed Akhter Hossain **Chairman**
Professor and Head
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

____

Narayan Ranjan Chakraborty **Internal Examiner**
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

____

Md. Tarek Habib **Internal Examiner**
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

____

Dr. Mohammad Shorif Uddin **External Examiner**
Professor
Department of Computer Science and Engineering
Jahangirnagar University

DECLARATION

I hereby declare that, this project has been done by us under the supervision of **Anup Majumder, Lecturer, Department of CSE** Daffodil International University. I also declare that neither this project nor any part of this project has been submitted elsewhere for award of any degree or diploma.

Supervised by:

____

Anup Majumder
Lecturer

Department of CSE

Daffodil International University

Co-Supervised by:

____

Masud Rabbani
Lecturer

Department of CSE

Daffodil International University

Submitted by:

____

Md. Jaowad Hasan

ID: 151-15-5495

Department of CSE

Daffodil International University

ACKNOWLEDGEMENT

First I express my heartiest thanks and gratefulness to almighty Allah for His divine blessing makes me possible to complete this project successfully.

I fell grateful to and wish my profound my indebtedness to **Anup Majumder, Lecturer,** Department of CSE, Daffodil International University, Dhaka. Deep Knowledge & keen interest of my supervisor in the field of Research in Image Processing influenced me to carry out this research. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior draft and correcting them at all stage have made it possible to complete this research.

I would like to express my heartiest gratitude to **Professor Dr. Syed Akhter Hossain**, Head**,** Department of CSE, and **Md. Tarek Habib,** Assistant Professor**,** Department of CSE, Daffodil International University, Dhaka and for their kind help to finish my research and also to other faculty member and the staff of CSE department of Daffodil International University.

I would like to thank my entire course mate in Daffodil International University, who took part in this discuss while completing the course work.

Finally, I must acknowledge with due respect the constant support and patients of my parents.

ABSTRACT

This project titled "Image Processing Based Fabric Detection". The purpose of this project is to detect fabric. Lack of knowledge about fabric can cause people to have lower quality or wrong type of fabric. Image processing is very complex but we are making progress in this sector. Many difficult problems are being solved by image processing. The utility of image processing in today's world is vastly accepted and used. Object or type detection is a very important part in image processing. In this research we have developed a method that can detect two of the most common types of fabrics – Cotton and Polyester using the available methods of classical image processing. A simple image takes from the smart phone can tell them the right type of fabric with high accuracy. Our method is trained with many images. The method uses decision tree based feature selection method to find important features of an image out of large feature space and SVM to classify image with the knowledge from the training data. We have obtained ~90% accuracy in successfully separating two above mentioned types of fabrics from cell phone images.

TABLE OF CONTENT

List of Figures

List of Tables

CHAPTER 1

INTRODUCTION

1.1 Introduction

From the moment we wake up and open our eyes, we are processing images. It is image processing that helps us to complete our works. For centuries people have to visually represent things through painting or drawings, now by taking pictures. Since the emergence of computers and particularly digital imaging, images now correspond to arrays of quantized values," said William Guicquero, a Ph.D. research engineer at Grenoble, France-based Leti, a technology research institute at CEA Tech and one of the world's largest organizations for applied research in microelectronics and nanotechnology. Image processing is very complex but we are making progress in this sector. Many difficult problems are being solved by image processing. We are making our daily work very easy and time consuming. There are many things that can be done by image processing. It is now used vastly everywhere. Detection is a very important part in image processing. In this research we have developed a method that can detect two of the most common type of fabrics – Cotton and Polyester using the available methods of classical image processing.

The detection of fabric is not a very popular application of image processing as of yet. However, for various purpose such as evaluating garments and quality of clothes, it can be useful to know right kind of fabrics that are used. An automated system that can detect a particular type of fabric and separate from another type can be helpful in many of our daily use. The availability of smart phones makes it easy for everyone to take pictures. Thus an accurate method of fabric detection from image taken by smart phones can be a useful application.

1.2 Motivation

Lack of knowledge about fabric can cause people to have lower quality of wrong type of fabric. A large industry is based on garments and the use of right kind of fabric is essential for clothes and apparels. However, a large portion of people cannot correctly distinguish different types of fabric since they have less exposer to the different kinds. An automated system that can detect different types of fabric for the user using an image which is readily available, thanks to the abundance of smart phones in regular people, is immensely helpful and useful for various purpose. This can be useful at a single user level as well as industry level where large amount of fabric is processed each day for garments. Currently available methods are not specifically made for the detection of fabrics. For this reason, the idea of proposing a tool that can detect fabric type from image was very promising. We have collected many images locally. We have applied necessary feature selection technique and classification method to detect the right kind of fabric. For all our experiment we have used one of the most popular and widely used language Python 2.7.

1.3 Objectives

The objective of this research to detect fabric which can help people to know about the fabric type. After identifying the fabric people will know whether this type of fabric were looking for or if this type of fabric he/she wanted or not. The proposed method is user friendly, and can be implemented in any system.

- To reduce fraud
- To detect right type of fabric in large industrial
- To help people if they are getting the right fabric or not

1.4 Expected Outcome

The method is expected to identify two of the most popular fabrics in use – Cotton and Polyester. A simple image takes from the smart phone can tell them the right type of fabric with high accuracy. My method is trained with many images. The method is trained with the pictures of Cotton and Polyester type of fabric. The method is trained with many images that's why I'm expecting high accuracy. The method works in a systematic way. The method uses important feature of an image and SVM to classify image with the knowledge from the training data.

Chapter 2 briefly explain the background of image processing the use of image processing in our daily life. This chapter also mentions different types of learning in Machine learning and explains different types of popular classifier that are used. Chapter 3 explains the feature selection method that is used in this work. The result analysis and discussion about the obtained results are explained in Chapter 4. Finally, in Chapter 6 contains the conclusion and a potential future direction of the work

1.5 Report Layout

Chapter 1: Introduction

In this chapter I have discussed about the introduction, motivation of the work, objectives and expected outcome of the research work and the report layout.

Chapter 2: Background

I discussed about the background circumstances of my work. I also delivered the literature review, research summery, scope of the problem and challenges of the system.

Chapter 3: Research Methodology

This chapter is all about the procedure used to build the system.

Chapter 4: Experimental results and discussion

In this chapter all the experimental result that has been achieved by the proposed system is discussed along with the performance analysis and a summary of the result is covered.

Chapter 5: Conclusions

This chapter contains the conclusion part and the ideas of implication of further study on this topic.

CHAPTER 2

BACKGROUND STUDY

2.1 Introduction

In this chapter, I discuss on several research work done by researchers in the area of image processing.

Image processing is a technique that acquires an image and analyze it, enhance or collect useful information from images and finally it output the result in an explainable or apprehensible format. The image may be analyzed to find patterns that aren't visible by the human eye. People can take decisions after getting the output, sometimes the decisions also can be made by the machine itself.

In this research Python 2.7 is used to train and analyze the data from numerous images of carrots. Python 2.7 is a platform where the digital image processing algorithms are implemented.

Image Processing Based Fabric Detection is a system that can describe a system what it is containing and what does it mean. In this work computer vision based system is used recognize fabric from numerous fabric images. It is a supervised learning process where fabric names are used to label the classes.

2.2 Literature Review

An image is an array, or a matrix, of square pixels arranged in columns and rows. Image processing is a method to perform some operations on an image, in order to get an enhanced image or to extract some useful information from it. It is a type of signal processing in which input is an image and output may be image or characteristics/features associated with that image. Nowadays, image processing is among rapidly growing technologies. It forms core research area within engineering and computer science disciplines too [1].
There are two types of methods used for image processing namely, analogue and digital image processing. Analogue image processing can be used for the hard copies like printouts and photographs.
Image analysts use various fundamentals of interpretation while using these visual techniques. Digital image processing techniques help in manipulation of the digital images

by using computers. The three general phases that all types of data have to undergo while using digital technique are pre-processing, enhancement, and display, information extraction [2]. Objective of Image Preprocessing is process an image so that the resulting image is more suitable than the original for a specific application. Also a preprocessing method that works well for one application may not be the best method for another application. Human relies very much on our visual system (eyes & brain) to collect visual information about our surrounding. Visual information refers to images and video. In the past, we need visual information mainly for survival. Nowadays, visual information is for survival as well as communication and entertainment. Equipment and software to capture, display, store and process images/video are getting cheaper and having better quality. More & more images/video are used. For example, images/video are common on Internet and mobile phones nowadays. Human visual system is highly non-linear [3]. Since we are working here with computer vision, Computer vision is a kind of automated watchdog, which uses both science and technology. Being a discipline from science, computer vision is related to theory for design of artificial systems that can acquire information from images. The image input may be of many formats, such as a video signal sequence, or multiple views from different cameras, or data input from a medical scanning machine. Examples of applications of computer vision include systems for controlling processes such as an industrial robot or an autonomous vehicle; for detecting events such as in visual surveillance or people counting; for organizing information such as for indexing databases of images and image sequences; for modeling objects or environments such as industrial inspection, medical image analysis or topographical modeling; for interaction such as the input to a device for interaction between a computing machine and human.

An image to be processed by computer must be represented using an appropriate discrete data structure, for example, a matrix. An image captured by sensor is exposed as a continuous function $f(x, y)$ is sampled into matrix with M rows and N columns. Image quantization assigns to each continuous sample an integer value – the continuous range of the image function $f(x, y)$ is split into k intervals. The finer the sampling and quantization, the better the approximate of the continuous image function $f(x, y)$ achieved.

Image function sampling poses two questions. Frist, the sampling period should be determined – this is the discrete between two neighboring sampling points in the image. Second, the geometric arrangement of sampling points should be set [4]. Some of the important applications of image processing in the field of science and technology include computer vision, remote sensing, feature extraction, face detection, forecasting, optical character recognition, finger-print detection, optical sorting, argument reality, microscope imaging, lane departure caution system, Non-photorealistic representation, medical image processing, and morphological imaging [5]. The image classification follows the steps as pre-processing, segmentation, feature extraction and classification. In the Classification system database is very important that contains predefined sample patterns of object under consideration that compare with the test object to classify it appropriate class. Image Classification is an important task in various fields such as biometry, remote sensing, and biomedical images. In a typical classification system image is captured by a camera and consequently processed. In Supervised classification, first of all training took place through known group of pixels. The trained classifier used to classify other images. The Unsupervised classification uses the properties of the pixels to group them and these groups are known as cluster and process is called clustering [6]. SVMs are set of related supervised learning methods used for classification and regression .They belong to a family of generalized linear classification. A special property of SVM is, SVM simultaneously minimize the empirical classification error and maximize the geometric margin. So SVM called Maximum Margin Classifiers. SVM is based on the Structural risk Minimization (SRM). SVM map input vector to ahigher dimensional space where a maximal separating hyper plane is constructed. Two parallel hyper planes are constructed on each side of the hyper plane that separate the data. The separating hyper plane is the hyper plane that maximize the distance between the two parallel hyper planes. An assumption is made that the larger the margin or distance between these parallel hyper planes the better the generalization error of the classifier will be [8]. SVM requires that each data instance is represented as a vector of real numbers. Hence, if there are categorical attributes, we first have to convert them into numeric data. We recommend using m numbers to represent an m-category attribute. Only one of the m numbers is one, and others are zero. For example, a three-category attribute such as {red, green, blue} can be represented as (0,0,1), (0,1,0), and (1,0,0). Our experience indicates that if the number of values in an attribute is not too large, this coding might be more stable than using a single number. Scaling before applying SVM is very important. In general, the RBF kernel is a reasonable first choice. This kernel nonlinearly maps samples into a higher dimensional space so it, unlike the linear kernel, can handle the case when the relation between class labels and attributes is nonlinear. Furthermore, the linear kernel is a special case of RBF Keerthi and Lin (2003) since the linear kernel with a penalty parameter $\tilde{C}$ has the same performance as the RBF kernel with some parameters (C, γ). In addition, the sigmoid kernel behaves like RBF for certain parameters (Lin and Lin, 2003).The second reason is the number of hyper parameters which influences the complexity of model selection. The polynomial kernel has more hyper parameters than the RBF kernel. Finally, the RBF kernel has fewer numerical difficulties [9]. Classification is one of the most important tasks for different application such as text categorization, tone recognition, image classification, micro-array gene expression, proteins structure predictions, data Classification etc. Most of the existing supervised classification methods are based on traditional statistics, which can provide ideal results when sample size is tending to infinity. However, only finite samples can be acquired in practice. In this paper, a novel learning method, Support Vector Machine (SVM), is applied on different data (Diabetes data, Heart Data, Satellite Data and Shuttle data) which have two or multiclass.SVM, a powerful machine method developed from statistical learning and has made significant achievement in some field. Introduced in the early 90's, they led to an explosion of interest in machine learning. The foundations of SVM have been developed by Vapnik and are gaining popularity in field of machine learning due to many attractive features and promising empirical performance. SVM method does not suffer the limitations of data dimensionality and limited samples [10]. Support Vector machines realize the ideas outlined above. To see why, we need specify two things: the hypothesis spaces used by SVM, and the loss functions used. The folklore view of SVM is that they find an "optimal" hyper plane as the solution to the learning problem. The simplest formulation of SVM is the linear one, where the hyper plane lies on the space of the input data x. In this case the hypothesis space is a subset of all hyper planes of the form: $f(x) = w \cdot x + b$. An application of SVM regression was discussed in (Fernandez, 1999). The problem was time series prediction. The approach taken was the use of SVM regression to model the dynamics of the time series and subsequently predict future values of the series using the constructed model. Instead of using the standard SVM regression formulation described above, a variation developed in (Scholkopf et al., 1998) was used. Using this variation the ε parameter of the SVM regression loss function (see above) is automatically estimated. Furthermore, (Fernandez, 1999) used an approach to learning which is different from the standard one: instead of developing one global regression model from all the available training data, (Fernandez, 1999) develops a number of SVM regression models, each one trained using only part of the initial training data. The idea, which has been suggested in (Bottou and Vapnik, 1992), is to split the initial training data set into parts, each part consisting only of training data that are close to each other (in a Euclidean distance sense). Then a "local" SVM is trained for each subset of the data. The claim in (Bottou and Vapnik, 1992) is that such an approach can lead to a number of simple (low complexity, in the SLT sense outlined above) learning machines, instead of a single machine that is required to fit all data. In (Fernandez, 1999) each of the individual SVM machines had its ε parameter estimated independently. The ε parameter of the SVM loss function is known to be related to the noise of the data (Pontil et al., 1998). So, in a sense, the approach of (Fernandez, 1999) leads to local SVMs each having a ε parameter that depends on the noise of the data in particular regions of the space (instead of a single ε that needs to "model" the noise of all the data). The experiments described in (Fernandez, 1999) show that training many local SVMs instead of one global learning machine leads to significant improvements in performance. In fact, this was also the finding of (Bottou and Vapnik, 1992) who first showed experiments with local learning machines [11]. Machine learning, a branch of artificial intelligence, is ascientific discipline concerned with the design and development of algorithms that allow computers to evolve behaviors based on empirical data, such as from sensor data or databases. A major focus of machine learning research is to automatically learn to recognize complex patterns and make intelligent decisions based on data. The SVM is already known as the best learning algorithm for binary classification. The SVM, originally a type of pattern classifier based on a statistical learning technique for classification and regression with a variety of kernel functions, has been successfully applied to a number of pattern recognition applications. Recently, it has also been applied to information security for intrusion detection. Support Vector Machine has become one of the popular techniques for anomaly intrusion detection due to their good generalization nature and the ability to overcome the curse of dimensionality .Another positive aspect of SVM is that it isuseful for finding a global minimum of the actual risk usingstructural risk minimization, since it can generalize well with kernel tricks even in high-dimensional spaces under littletraining sample conditions. A major focus of machine learning research is to automatically learn to recognize complex patterns and make intelligent decisions based on data. The SVM can select appropriate setup parameters because it does not depend on traditional empirical risk such as neural networks [12]. One of the main advantage of using SVM for IDS is its speed, as the capabilityof detecting intrusions in real-time is very important. SVMscan learn a larger set of patterns and be able to scale better, because the classification complexity does not depend on the dimensionality of the feature space. SVMs also have the ability to update the training patterns dynamical whenever there is a new pattern during classification [13].

2.3 Research Summary

In this research so many images of fabric are collected from various places. Then the images are pre-processed for further processing. Features of the images are evaluated and extracted from the images by tree based classifier. Dataset is prepared and trained along with specific label for each class of the fabric which is done in Python 2.7 through the calculation of features. When all the processing is done our system is ready to use. Test image is captured and given as input after preprocessing. The features are extracted and the selected classifier image is compared with the previous dataset. SVM classifier is used to identify the class name or the disease name. Finally the output of the system is shown.

2.4 Scope of the Problem

The proposed framework can help people as well as industries to have benefits of modern technology to prevent improper loss of their profit. Although the whole process like capturing images, entering the image as input and the implementation of the application may be a fact to a non-professional user who do not have any knowledge of this type of system. But this is a proposed framework which includes the main idea of steps how to process the data and how the algorithms should be implemented. It can be implemented in any kind of platform regardless of choice. Using this approach mobile applications or online based web applications can be developed to reach the people easily.

2.5 Challenges

To build such a system it needs images of fabric to be trained first and then some more images are needed for the testing purpose. Images were collected from many places.

The big challenge with this approach is the quality of the images. Images of low resolution cannot be processed easily. And also colleting images was very challenging.

CHAPTER 3

RESEARCH METHODOLOGY

3.1 Introduction

In this chapter I am going to explain the feature selection method I have applied for my classification problem. Feature selection is an important part of any machine learning method. Working with reduced number of important feature not only improves the classification performance, moreover, it helps reduces over fitting the model towards dataset specific problem and builds a more general framework.

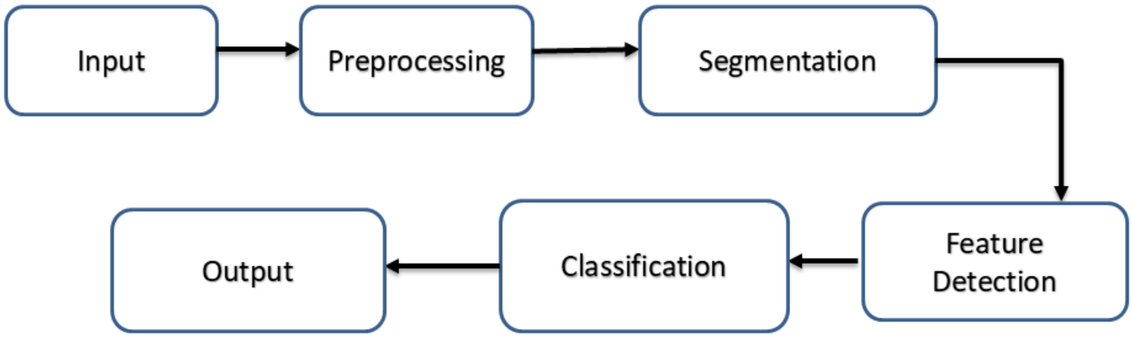


Figure 3.1: Research Methodology

As shown in figure 3.1 after inputting an image the image is processed and segmented then the features are getting detected and after that classifier do the classification and then give an output.

3.1.1 What is feature selection?

Feature selection is the process of choosing selective attributes that are relevant to the predictive model. The general approach of feature selection reduces the number of initial attributes the dataset has. However, feature selection is not just reducing the dimension of the dataset which creates a new combination of attributes, but selection of important attributes and rejection of relatively less useful attributes. Figure 3.2 shows the general position of feature selection methods in a classification problem pipeline.

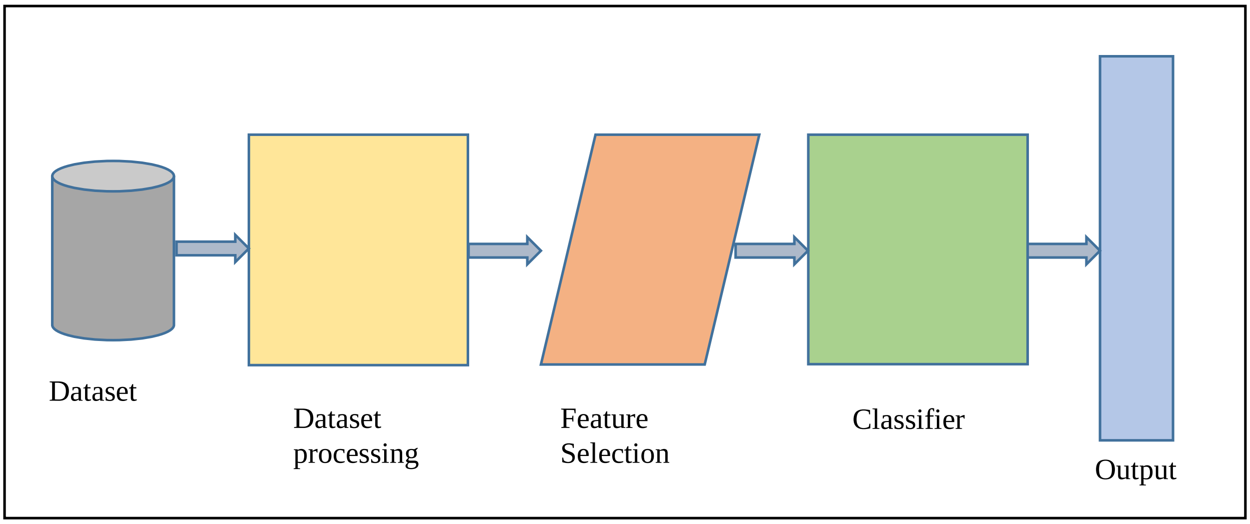


Figure 3.2: Feature selection in the classification

As shown in the figure 3.2, a feature selection method takes in processed data that are suitable for the problem in hand and selects attributes that will help classifier better perform over the dataset. The classifier then outputs the results.

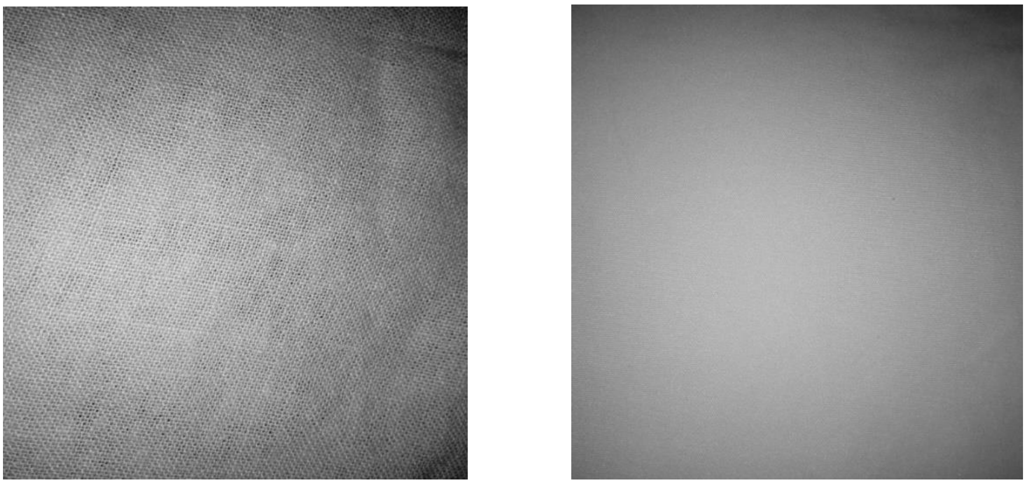
I have collected images of two different types of fabrics – Cotton and Polyester. The images of the fabrics were taken with smart phone cameras. The details of the working dataset is given in Chapter 4.

3.1.2 Feature selection method

The structure of the two type of fabrics are different. As you can see in Figure 3.3 the difference in their making is distinguishable.

The images shown in Figure 3.3 is in grayscale. The images were taken as colored images and converted into grayscale so that the range in the pixel remains [0, 255]. This work is fundamentally different from classical image processing such as object detection, edge detection of activity detection. The purpose of this work is to identify two different type of fabrics and classify them accurately. That is why the feature selection approach for this work is also different than the classical image processing methods.

It is noteworthy to mention that the dataset we are using in this work is collected and build by myself. As a result, the quality of the data has a scope of improvement. The angle and lighting of the capture plays an important role for an image.



Type: Cotton Type: Polyester

Figure 3.3: Images of two types of the fabrics

This work contains very simple capture with almost constant lighting. In the real world the images might not be available in this fashion. To get the fabric detected accurately, image should be segmented first.

3.2 Methods and steps

In this experiment the raw image goes through several preprocessing before applying the feature selection on the image data. Below I am briefly describing each step to process the dataset.

3.2.1 Converting image to numerical matrix

I have collected raw colored image with smart phones. To process the image first I need to convert the colored image to grayscale images. Each image is read by imread function available in OpenCV package in python. The parameter passed in the function converts the image to a gray scale image and stores the image in a predefined resized $n \times n$ matrix. For this experiment we have selected $n = 128$.

Thus each image is converted into a 128 x 128 numerical matrix ranging [0, 255]. The image is resized to save the storage space as well as preventing the feature selection method from overwhelming feature space.

3.2.2 Reshaping matrix to vector

Once all the images are converted into 128 x 128 matrices, I reshaped each matrix into a vector. The size of the vector is 128 x 128 = 16,384. For m images, the training is then reshaped from $m \times 128 \times 128$, 3-dimensional matrices to $m \times 16384$, 2-dimensional matrix. The row of this matrix represent each sample and the column of the matrix represents features across m samples. Reshaping the dataset from 3-dimensional to 2-dimensional matrix gives us more control over dataset to perform feature selection and classification.

3.2.3 Scaling the data

After reshaping the dataset, I standardize the data by scaling the data. For scaling the dataset I have used sklearn.preprocessing.MinMaxScaler. This step transforms the feature by scaling each feature to a given range. For each column first the method finds the standard deviation (std.) and scales the data. The scaling can be defined by

$$std\ X_{col} = \frac{X_i - X_{min}}{X_{max} - X_{min}}$$

$$X_{scaled} = std\ X_{col} - (\max - \min) + min$$

Where max and min are the range of values in specific feature column. This transformation is necessary for the method so that the feature selection and the classifier is not biased towards extreme data in the dataset.

3.2.4 Feature selection

This work uses a tree based estimator to find important set of features from the dataset. For the feature selection we have used off-the-shelf method named ExtraTreesClassifier available in sklearn.feature_selection imported from SelectFromModel.

This readily available class implements a meta estimator that fits many randomized decision tree on various subset of samples from the dataset and uses averaging to improve the predictive performance and balancing the over-fitting. The feature selection method uses a parameter called n_estimator which is set to 50 for our experiment.

This number defines the number of tree in the forest while making decisions over many decision trees. Figure 3.4 shows random trees from the forest build Using subset of samples available in the dataset.

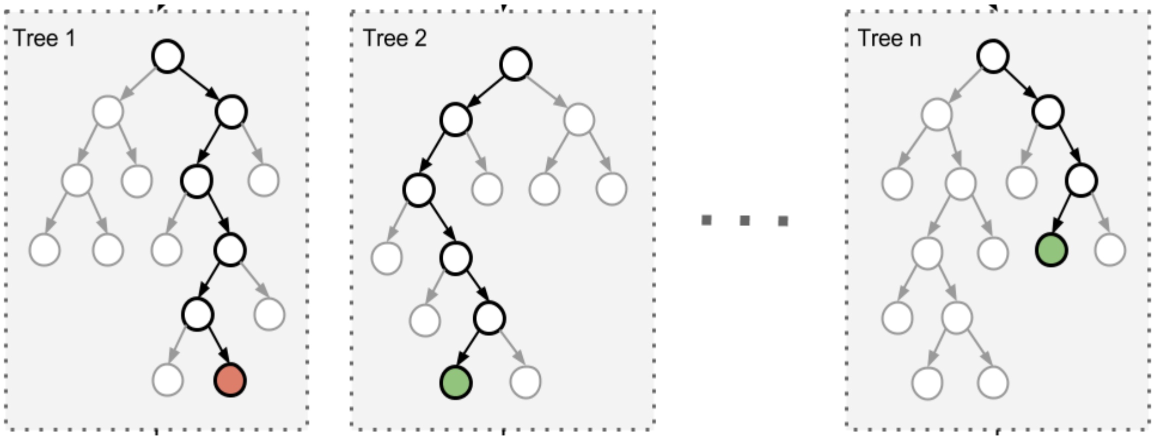


Figure 3.4: Randomized tree in the forest. Image credit [7]

The feature selection method provides us a list of indices that are important for the classifier to classify. I keep those indices and format out dataset according to the indices returned by the feature selection method.

Once the data is correctly formatted with the important features, the training and testing data is ready for the classifier. The next chapter explains the results of my experiment.

CHAPTER 4

Experimental Results and Discussion

4.1 Introduction

This chapter illustrates the results I have obtained from the method I have built. In this section I will briefly explain about our dataset, our experimental setup and the results that I have obtained. Lastly I am going to discuss the results and future direction of the work.

4.2 Dataset

As I have mentioned in the previous chapter, I have collected our own dataset by capturing images of Cotton and Polyester fabric available locally. I have collected over 500 images of the two types of the fabrics. I have balanced the dataset by assigning training and testing set with equal number of two types of the data. Balancing the dataset was intentional so that the feature selection and classification is balanced and I have a general classifier model without bias. Table 4.1 shows a summary of the dataset used in the experiment. The experiment uses 67% (335) of the total images for feature selection and to train the model.

Table 4.1: Basic statistics of the dataset used in the experiment

Type of Fabric	**Training data (67%)**	**Testing Data (33%)**
Cotton	167	83
Polyester	168	82
Total	335	165

The rest of the 33% (165) of the images I have used to test the model. Here in training data system is taking 167 images of cotton and 168 images of polyester in total of 335 images of cotton and polyester. and in testing the system is taking 83 images of cotton and 82 images of polyester in total of 165 image for testing. The experiment is performed multiple times and each experiment uses randomly selected training and testing data at each iteration. The results are explained in following sections.

4.3 Experimental Setup

For this experiment I have used python 2.7. For reading image I have used OpenCV. I have used sklearn package for randomly splitting the dataset into training and testing dataset. In the preprocessing step, the data is scaled with MinMax method from the sklearn.preprocess package. ExtraTreesClassifier is used from the sklearn.ensemble for feature selection step. For classifying the data, I have used SVM available in sklearn.

4.4 Parameters

This experiment uses various parameters at feature selection step and classification step. At the feature selection step I set the number of estimator to 50 for extra trees classifier while evaluating feature importance. The other two parameter used at the classification step while the data is being classified by SVM. Two important parameter for SVM performance is C and gamma. C defines the weight of how much samples inside the margin contribute to the overall error. Gamma can be inferred as the inverse of the radius of the influence of samples selected as support vectors by SVM. I have performed a grid search of these two parameters. I searched for $\mathrm{C} = [64, 128, 256, 512]$ and $\mathrm{gamma} = [0.0001, 0.001, 0.01, 0.1]$. The grid search performed 10-fold cross validation to find the best parameter over 4×4 parameter grid of C and gamma. After 10-fold cross validation on the training dataset the best parameter was chosen as $\mathrm{C} = 64$ and $\mathrm{gamma} = 0.01$ with a 10-fold cross validation accuracy of 0.946. For both 10-fold cross validation process and classification process I have used the RBF kernel in SVM classifier.

4.5 Variation in number of features selected

I have performed the experiment 10 times to obtain a generalized performance of the model. At each iteration the model selects different number of features. Because every time the system is taking different type of data set. The images are given in a folder and in every experiment the system selects its training data and testing data randomly that’s why the number of features are different every time. Figure 4.1 shows the variation in selection of the number of features.

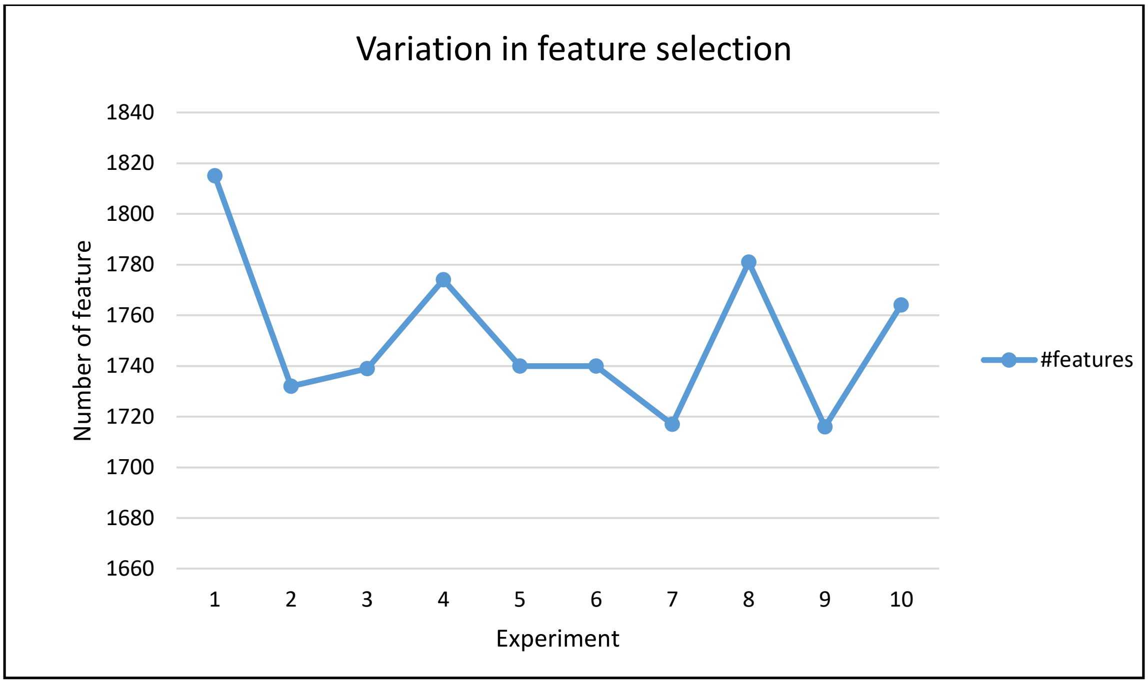


Figure 4.1: Variation in number of feature selected in each experiment.

Figure 4.1 shows the change in the number of features at each iteration of the experiment. The highest number of feature was selected at experiment 1 where the number of feature is 1,815 while the least number of feature is selected at experiment 9 where the number of feature is 1,716. One of the obvious reason of varying number of features is the random selection of training data at each experiment. I will analyze how the variation in the number of feature affects the performance of our model.

4.6 Classification performance

To measure the classification performance, I have used the accuracy, sensitivity and specificity as our metrics. Each of the metrics is defined by

$$sensitivity = \frac{TP}{TP + FN}$$

$$specificity = \frac{TN}{TN + FP}$$

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Where TP = True Positive (correctly identified), FP = False Positive (incorrectly identified), TN = True Negative (correctly rejected) and FN = False Negative (incorrectly rejected). Table 4.2 shows the performance of our model with the above mentioned metrics.

Table 4.2: Performance evaluation of the method on 10 experiment

Iteration	**#features**	**Accuracy**	**Sensitivity**	**Specificity**
1	1815	0.896969697	0.867346939	0.828947368
2	1732	0.909090909	0.911111111	0.894736842
3	1739	0.903030303	0.892473118	0.868421053
4	1774	0.890909091	0.88172043	0.855263158
5	1740	0.878787879	0.855670103	0.815789474
6	1740	0.890909091	0.873684211	0.842105263
7	1717	0.903030303	0.87628866	0.842105263
8	1781	0.915151515	0.912087912	0.894736842
9	1716	0.903030303	0.884210526	0.855263158
10	1764	0.890909091	0.88172043	0.855263158
Average	**1752**	**0.898181818**	**0.883631344**	**0.855263158**

As we can see from Table 4.3, the model performs well on all iteration. The maximum accuracy was obtained at iteration 8 with an accuracy of 0.92 and lowest accuracy was obtained at experiment 5 with an accuracy of 0.87. The average accuracy was 0.89 with and standard deviation of 1%. Similarly, the average sensitivity is 0.88 with a standard deviation of 1% and average specificity of 0.85 with 2% standard deviation.

Table 4.3 shows the performance by computing the real numbers that are obtained from the model. The numbers represent how many times the model correctly identified/rejected and incorrectly identify/rejected a test sample.

Table 4.3: Correctly and incorrectly identified sample count

Iteration	Criteria	Cotton	Polyester
Experiment 1	Correctly Identified	85	63
	Wrongly Identified	13	4
Experiment 2	Correctly Identified	82	68
	Wrongly Identified	8	7
Experiment 3	Correctly Identified	83	66
	Wrongly Identified	10	6
Experiment 4	Correctly Identified	82	65
	Wrongly Identified	11	7
Experiment 5	Correctly Identified	83	62
	Wrongly Identified	14	6
Experiment 6	Correctly Identified	83	64
	Wrongly Identified	12	6
Experiment 7	Correctly Identified	85	64
	Wrongly Identified	12	4
Experiment 8	Correctly Identified	83	68
	Wrongly Identified	8	6
Experiment 9	Correctly Identified	84	65
	Wrongly Identified	11	5
Experiment 10	Correctly Identified	82	65
	Wrongly Identified	11	7

From Table 4.3 we can observe that the number of times out model correctly identifies Cotton is very high compared to the correct identification of Polyester. Moreover, the number of incorrectly identifying Cotton is also higher than Polyester which leads to the conclusion that the model is slightly biased towards the images classified as Cottons.

4.7 Effect of varying feature size over performance

As I have explained in the previous section that the number of feature selected by the feature selection method is varying based on the randomly selected training data. Here in this section I am evaluating the change of the number of features over the variation in the performance. From Figure 4.2 we can observe that the variation in the size of the feature does affect the performance. The best performance is obtained with the feature size 1,781 and the worst very performance was obtained with 1,740 features.

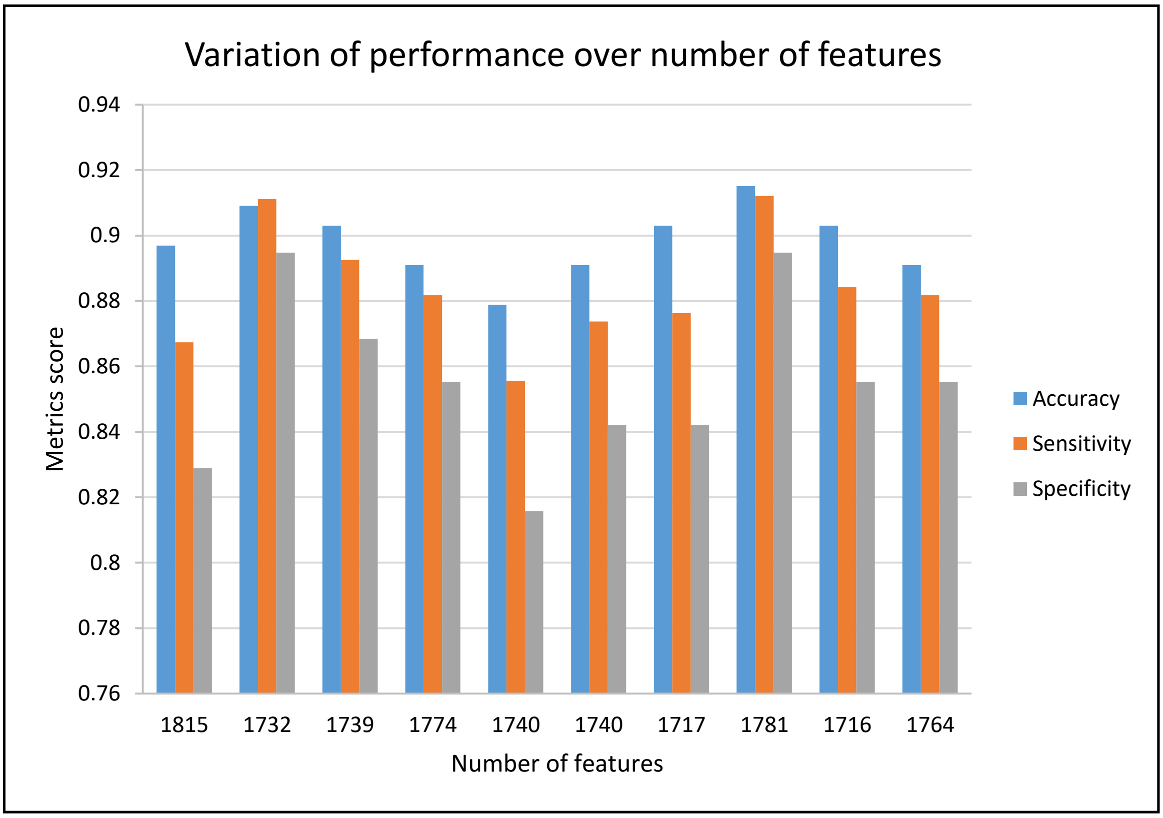


Figure 4.2: Variation of performance for varying number of feature at different experiment.

However, it cannot be concluded that the performance is dependent on the number of feature since neither best or worst performance shows no direct correlation with the highest or lowest number of features.

As a result, this can be concluded the performance is strictly depended on the quality of the features and the features extracted from the images selected for the experiment.

The overall performance of our model is satisfactory. I am able to correctly classify the two types of fabric with an average accuracy of 89% with high sensitivity and specificity. Although the performance varies over different experiment, but the overall evaluation of the metrics shows a satisfactory performance of the model.

CHAPTER 5

Conclusions and Implication for the Future Research

5.1 Conclusion

In this work I have proposed an image classification method that classify two types of fabrics from images. I have collected 500 images of Cotton and Polyester fabrics. To implement our method, I have preprocessed the data and applied feature selection method on the data to find important feature for the classification. I have chosen SVM as my classifier and decision tree based method as my feature selection method. The parameters for feature selection and classification method was optimized. I have implemented our work on python. As I have explained in the previous section, I have obtained very good classification accuracy high very high sensitivity and specificity.

5.2 Implication of further study

With the success of this work, I am planning to extend my work to multi-class classification problem, where my model will be able to classify more than two types of fabrics. I have already obtained images of one other type of fabric. Also, I am planning to tune my model so that the model can handle different kind of images such as fabric images from different angles and lighting. Also varying quality of images will be trained to build a generic fabric detection tool.

REFERENCES

[1] Golamreza Anbarjafari PhD, *Video lecture on digital image processing*, Available at: https://www.ut.ee/en/mooc/digital-image-processing, [last accessed: 20 November 2018].

[2] Dr. Romik Chatterjee, Vice President of Engineering, Graftek Imaging, Inc., *Image Processing Fundamentals*, , Graftek Imaging, Inc.

[3] Mohammed A. Hasan, *Introduction to Digital Image Processing*, Department of Electrical & Computer Engineering, University of Minnesota-Duluth

[4] A. Kumar and F. Shaik, *Image Processing in Diabetic Related Causes*, Forensic *and Medical Bioinformatics.*

[5] Shailendra Kumar Dewangan, Shailendra Kumar Dewangan, Assistant Professor, Department of Electronics & Instrumentation Engineering, Chhatrapati Shivaji Institute of Technology, Durg, Chhattisgarh, India, *International Journal of Computer Science & Engineering Technology (IJCSET)* . Institute of Technology, Durg, Chhattisgarh, India

[6] Rama Gaur, Jodhpur National University, Jodhpur, Rajasthan, India,Professor , Dr. V.S. Chouhan,Ph.D. Scholar (ECE), Head ECE department ,MBM Engineering college, Jodhpur, Rajasthan, India, *Classifiers in Image processing.* Jodhpur National University, Jodhpur, Rajasthan

[7] Decision Tree. Available at: https://blog.statsbot.co/ensemble-learning-d1dcd548e936 [last accessed: 20 November 2018].

[8] V. Vapnik. *The Nature of Statistical Learning Theory*. NY: Springer-Verlag. 1995.

[9] Chih-Wei Hsu, Chih-Chung Chang, and Chih-Jen Lin, *A Practical Guide to Support Vector Classification*, Department of Computer Science, National Taiwan University, Taipei 106, Taiwan.

[10] DURGESH K. SRIVASTAVA, Ass. Prof., Department of CSE/IT, BRCM CET, Bahal, Bhiwani, Haryana, India-127028, LEKHA BHAMBHU, Ass. Prof, Department of CSE/IT, BRCM CET, Bahal, Bhiwani, Haryana, India-127028, *DATA CLASSIFICATION USING SUPPORT VECTOR MACHINE*, Journal of Theoretical and Applied Information Technology.

[11] Theodoros Evgeniou and Massimiliano Pontil, *WORKSHOP ON SUPPORT VECTOR MACHINES: THEORY AND APPLICATIONS*, Center for Biological and Computational Learning, and Artificial Intelligence Laboratory, MIT, E25-201, Cambridge, MA 02139, USA.

[12] Jayshree Jha, Department.of Computer Engineering, RamaraoAdik Institute of Technology, Navi Mumbai, India, Leena Ragha, Ph.D, Department Of Computer Engineering, RamaraoAdik Institute of Technology, Navi Mumbai, India, *Intrusion Detection System using Support Vector Machine*. , Department Of Computer Engineering, RamaraoAdik Institute of Technology, Navi Mumbai, India

[13] SandyaPeddabachigari, Ajith Abraham, CrinaGrosan, Johanson Thomas (2005). *Modeling Intrusion Detection, Systems using Hybrid Intelligent Systems*. Journal of Network and Computer Applications.