Cotton Contaminants Detection and Classification using HSI and YCbCr Model

Aditi Sachar[1], Sugandha Arora[2]

[1, 2]*Department of Computer Science & Engineering, PTU, SVIET, Banur, Patiala, India*

Abstract—*Contaminants present in cotton have serious impact on the quality of cotton fiber mainly in terms of fineness and strength of cotton fiber. The manual removal of cotton contaminants requires lot of manpower and is a time consuming process. So, in textile industries automatic cotton contamination detection is required as manual removal of contaminants has low efficiency. The detection of cotton contaminants is particularly challenging due to the large number of contaminant classes which are characterized by their vagueness and ambiguity based on their characteristics and due to their unpredictable size, shape, material and position as some of the contaminants get inside the cotton fiber layer and become invisible and some are of same color as cotton fiber. Digital image processing algorithm based on machine vision provides accurate and efficient detection of contaminants in real-time. In this research, a comparison of HSI and YCbCr color model is presented in detecting and classifying contaminants. For detecting different types of contaminants, automatic thresholding is used. After detection, naive bayes classifier is used for differentiating different types of contaminants that worked on the principle of feature extracted from different contaminants from the cotton fiber. Shape descriptors like extent, solidity, area, orientation are used as features to distinguish contaminant classes.*

Keywords—*Cotton contaminants, binarisation, thresholding, feature extraction, naive bayes classifier.*

I.   INTRODUCTION

Cotton is the most popular fabric in the world. It is a natural fiber that is harvested from the cotton plant and is used to make many fabric types. The quality of cotton fiber is degrading due to the presence of contaminants like plastic film, nylon straps, jute, dry cotton, bird feather, paper and various foreign fibers like silk, nylon, polypropylene etc.

Contamination of raw cotton can take place at every step i.e. from the farm picking to the ginning stage. Since cotton is picked manually by rural women so human hair, contamination caused by cloth pieces and fabric sheet are the biggest cause of cotton contamination. In addition foreign fibers including cloth strips, plastic film, jute, hair, polypropylene twine and rubber are serious threat to the textile and cotton industry. Such contaminants have effect on cotton grade and can cause color spots in fabric, thus reduce the textile value as well.

The blow room machinery [1] plays an important part in reducing the quantity of foreign particles in cotton but even this process can not remove all the contaminants and the leftover embedded pieces of contaminants can affect the quality of yarn and its value and the contaminants such as stones, metal pieces, and etc. causes disturbance to material flow especially affect production as well as quality of the machinery. In the manual process of cotton contamination detection, it is difficult to detect and classify the contaminants due to their unpredictable size, shape, material and position. So, automated instruments for detecting and removing foreign fibers in cotton are now developing to provide high performance and accuracy.

The different categories of cotton contaminants are shown in figure.1.
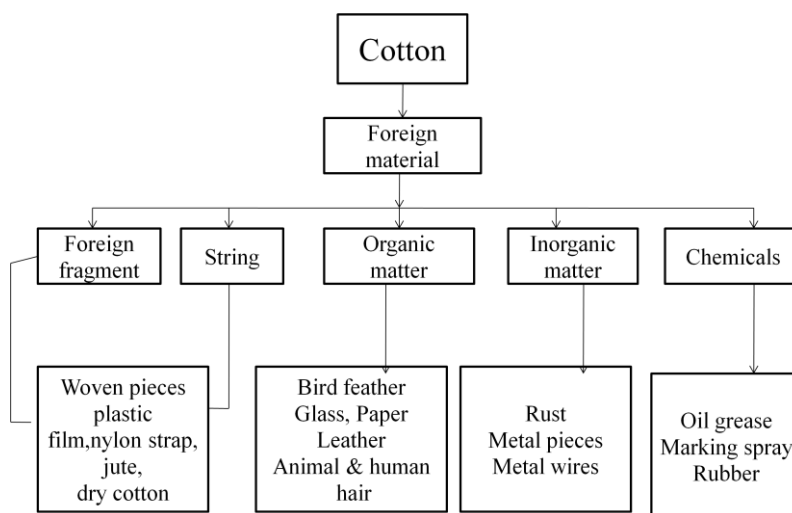

Figure 1. Types of contaminants.

The goal of this research is to develop real time algorithm for contaminant detection and classification. The objectives of this paper are:

i)   Detection of various cotton contaminants.
ii) Identification of features to recognize contaminants in cotton sample.
iii) Classification of contaminants as nylon, bark, metal, stones etc.
iv) Comparison of mean square error and time taken to detect and classify contaminants for HSI and YCbCr model.

II.   RELATED WORK

In recent years, automated vision systems are now used in textile industries. Zhenwei Su [2], Liwei Zhang [3], built real-time automated visual inspection systems for contaminant removal from wool. Boshra D. Farah [4] built AVI systems for inspection or removal of contaminants in cotton. Bidan Li [5] designed a machine vision system for detecting foreign fibers in lint.

Pai [6] proposed an algorithm that could detect and classify different types of contaminants via x-ray microtomographic image analysis. This technique could be used with fuzzy-logic-based classification scheme to create a highly accurate contaminant analysis tool.

Mingxiao Ding et al. [7] used the texture features to construct gray level co-occurrence matrix algorithm to detect the sharp contrast objects.

Cheng Liang Zhang et al. [8] proposed an approach for detecting contaminants using wavelets. This paper decomposed 2-D image signal into multi-layer wavelet by using wavelet packet 2-D. Experimental results show that the two-dimensional wavelet packet tools have very strong function in detecting foreign cotton fibers image.

Cheng Liang Zhang et al. [9] proposed an approach for detecting contaminants based on YCbCr color space. The advantage is that it can conduct various advanced algorithms to gray image from luminance, meanwhile, also can perform color detection to most colored foreign fibers and extract the chrominance information directly. Pooja Mehta & Naresh Mehta [10] presented a paper describing comparison between HIS model and YCbCr model and found that HSI model is better than the YCbCr model as YCbCr model was unable to distinguish the white colored contaminant from that of standard cotton whereas it was possible to detect the white fiber from the cotton in HSI model.

III.   PROPOSED APPROACH

3.1. Algorithm of Methodology for HSI & YCbCr model:
STEP I: Read the input RGB image having size 256* 256 pixels.
STEPII: Apply color transformation from RGB to HSI image & RGB to YCbCr image using the transformation formulae.
STEP III: Apply binarisation to convert the HSI and YCbCr images into binary images.
STEP IV: Convert the output of step III into gray scale image.
STEP V: Read the output of step IV and apply thresholding. Find the minimum and maximum threshold value for each contaminant.
STEP VI: Apply feature extraction to each output image of step V. Save the results into feature.mat file.
STEP VI: Load the feature.mat file into naive bayes classifier and calculate the accuracy as output for the system.

Figure below shows the basic steps followed in the methodology:

Figure 2. Work approach

IV.   SELECTION OF COLOR SPACE

A. RGB Color Space

RGB color space is the most fundamental and commonly used color space of image processing. Color information initially collected by image acquisition devices is RGB value, which is also finally used by color display devices. RGB model uses three basic components values of R, B and G to represent color. In this system, any color calculated is all within the RGB colorized cube. However, RGB color space has great shortcomings, the main one of which is that it is not intuitionistic, so it is hard for us to know color's cognitive attributes expressed by a value from its RGB value. Then, RGB color space is one of the most uneven color spaces, as the visual difference between two colors can not be expressed as the distance between two color points. In addition, the correlation between RGB is much high, and RGB space is sensitive to noise in low intensity area.

B. HSI Color Space

HSI means hue, saturation and intensity. HSI model decouples the intensity component from the color carrying information in a color image. As a result, it is an ideal tool for developing image processing algorithms based on color descriptions that are natural and intuitive to humans. As a result, it is an ideal tool for developing image processing algorithms based on color descriptions that are natural and intuitive to humans.

C. YCbCr Color Space

The detection algorithm of cotton foreign fibers in this paper compares HSI and YCbCr color model. The advantage of YCbCr model is that it can process luminance and chrominance information separately by excavating the useful information of the original image as more as possible. The original image is in the form of RGB, so it is necessary to transform color space. There are many color spaces, where the luminance and chrominance components are separated, such as YCbCr, HSV and Lab, etc. The paper adopts YCbCr color space. We transfer the pixel values of RGB space into luminance Y, chrominance of blue Cb and chrominance of red Cr in the YCbCr space.

The conversion formula used is:
$Y = 16 + (65.481\,R + 128.553\,G + 24.966\,B)$
$Cb = 128 + (-37.797\,R - 74.203\,G + 112.0\,B)$
$Cr = 128 + (112.0\,R - 93.786\,G - 18.214\,B)$

V.   EXPERIMENTS AND ANALYSIS

D. Detection of contaminants

Different types of contaminants namely bark, stones, plastic film, polypropylene twine, hair, leaves were selected for the experiments. Adequate samples of each contaminant were prepared and sample of pure contaminant was also prepared for detection.

In the experiment, images of cotton having contaminant were first acquired and then they were passed to the image processing algorithm for conversion from acquired RGB image to HSI and YCbCr images respectively. Figure 3(b) - (g) shows the output of various images processing algorithm followed in this research. Then each image is binarised to a particular threshold value to separate out the background from the contaminants. For each contaminant type, the range of thresholding was found out as shown in the table below:

TABLE 1. THRESHOLD RANGE FOR DIFFERENT CONTAMINANTS

Contaminants	Threshold values	
	Min Threshold value	Max threshold values
Leaf	15	115
Bark	10	103
Nylon	05	105
Hair	05	62
Stones	20	100

The above threshold values show an overlapping range between the different contaminants. Due to this, it becomes difficult to correctly classify different contaminants into different classes.

Binarisation after thresholding detects the different contaminants and after that features of different contaminants were extracted that were fed into the naïve bayes classifier that classifies the contaminants into different classes. Four classes of contaminants were passes to naive bayes classifier. These four classes included nylon, hair, bark and leaf.
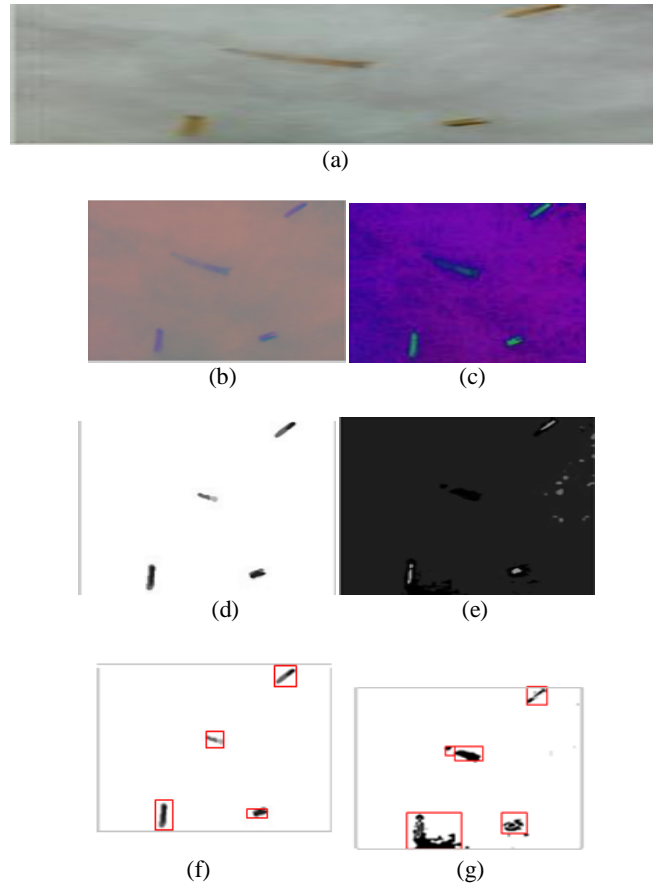
(a)

(b)   (c)

(d)   (e)

(f)   (g)

Figure 3. (a) Original image of cotton having contaminants.   (b) Transformed YCbCr image, (c) transformed HSI image, (d) binarised YCbCr image,   (e) binarised HSI image, (f)Thresholded YCbCr image showing different contaminants,   (g) Thresholded HSI image showing different contaminants.

E. Feature Extraction

The paper adopts the image processing method based on shape features which includes area, perimeter, solidity, extent, orientation, eccentricity, bounding box area and convex area. Some features were characteristics of the sample domain and some features such as area and perimeter could be easily used by human for separating the objects.

The image analysis software provided the following set of features:

  i.   **Area A**: Area of objects in a binary image, the number of nonzero pixels in an object.
  ii.   **Perimeter *P***: Perimeter of objects of a binary image, the length of the smoothest boundary in pixels.
  iii. **Convex Area**: Area of a 32-sided irregular polygon bounding the object.
  iv.   **Bounding box area**: Area of a rectangle circumscribing the object with sides parallel to image edges.
  v.   **Solidity**: Ratio of Area to Convex Area.
  vi.   **Extent:** Ratio of Area to Bounding Box Area.
  vii. **Eccentricity:** The eccentricity is the ratio of the distance between the foci of the ellipse and its major axis length. The value is between 0 and 1.

The feature extraction algorithm finds the numerical value of each feature for both HSI images and YCbCr images. Table 2 shows a set of values of features extracted for four contaminants taken for classification.

TABLE 2. FEATURE MEASUREMENTS

Extracted Features	Contaminants			
	Bark	Nylon	Hair	Leaf
Area (A)	1317	466	232	1028
Perimeter (P)	247	233	82	145
A/P	5.5	1.99	2.8	7.08
Extent	0.45	0.11	0.2	0.5
Eccentricity	0.16	0.43	0.23	0.6
Solidity	0.63	0.34	0.29	0.8

F. Classification methods

In this section, classification method based on naive bayes classifier is presented. Different trash types were placed on the cotton and used to collect the training data. Each acquired image is converted to HSI and YCbCr color space. Thresholding is used to obtain the threshold for separating the non-lint material from lint. The contaminants are labeled as blobs for the collection of data. Area, perimeter, ratio of area and perimeter, extent, eccentricity and solidity are computed for all the contaminants in the binary image.

- *Naive bayes classifier* [11]: A naive Bayes classifier assumes that the presence (or absence) of a particular feature of a class is unrelated to the presence (or absence) of any other feature, given the class variable. Depending on the precise nature of the probability model, naive Bayes classifiers can be trained very efficiently in a supervised learning setting. In many practical applications, parameter estimation for naive Bayes models uses the method of maximum likelihood.

- In this work the target vector was encoded by specifying the type of contaminant. For example, bark is encoded as 1, hair as 2, leaf as 3 and nylon as 4. Table 3.1 displays the contaminant classes and their corresponding target patterns.

TABLE. 3. TARGET PATTERN ENCODING

S.No	Contaminant Class	Target Pattern
1	Bark	1
2	Hair	2
3	Leaf	3
4	Nylon	4

The output of the naive bayes classifier is in the form of mean square error and confusion matrix. More the mean square error is close to zero better is the method. In this research, the values of mean square error is compared to find out which color space among HSI and YCbCr is better in classifying the contaminants.

- *Confusion matrix*: A confusion matrix is a specific table layout that allows visualization of the performance of an algorithm, typically a supervised learning one. Each column of the matrix represents the instances in a predicted class, while each row represents the instances in an actual class. The name stems from the fact that it makes it easy to see if the system is confusing two classes (i.e. commonly mislabeling one as another).

VI.   RESULTS

The result of detection by HSI and YCbCr is compared on the basis of time taken to perform the transformation, binarisation, thresholding and feature extraction. The graph below shows the total time taken to perform detection in both HSI and YCbCr color space.
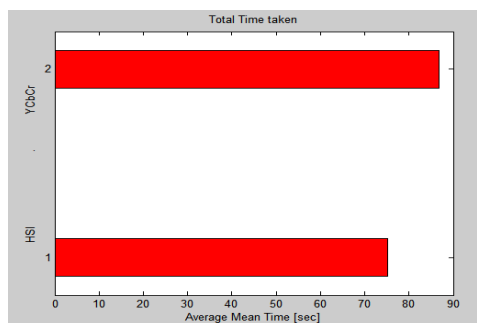

Figure 4. Graph showing total time taken in both color spaces

The above graph shows the total time taken for each color space. The time taken in HIS color space is 76.5 sec and in case of YCbCr is 88.7 sec. The time gap between the color spaces shows that HSI color space is better in detecting and classifying the contaminants as compared to YCbCr color space.

For identifying the contaminants in cotton,we need to design a pattern matching algorithm in such a manner that it works principally on the nature of the data set of the features of the contaminants.

For this, the data set of the features of the contaminants are first extracted and found that the data was not linearly separable as it was having overlapping threshold values.Since, the intersection of the classes within the data set of contaminants doesnot result in Φ.Therefore, the challenge is to find a classfier which will have right accuracy to identify the decision boundary.

For any classifier to have high accuracy, we need to design a discriminant function in such a manner that the discriminant is highly significant and at the same time it is having good speed for classification.

This is achieved by taking the mean of error rate which is the measure of errors. Therefore our attempt is to identify a classifier which will have minimum square error.To achieve best results ,the data set of contaminants was divided into four parts having labeled samples.The 70% of data was given for training classifier and 15% each was given for training and validation.We have used tenfold validation.

Interpretation of results:
Based on 70% samples collected from the contaminated data set, training was done. After running the training , the following result is obtained in case of HSI color model.
cMat2 =

```
60  158    0   32
 0  214    0   36
 0  110  109   31
 0  136   23   91
```
And the following result is obtained in case of YCbCr color model.
cMat2 =

```
92    0   64   94
21    0  146   83
 0    0  204   46
 0    0   69  181
```

It shows the classification and mis-classification done by the classifier.The diagonal values show the correct classification done by the classifier.
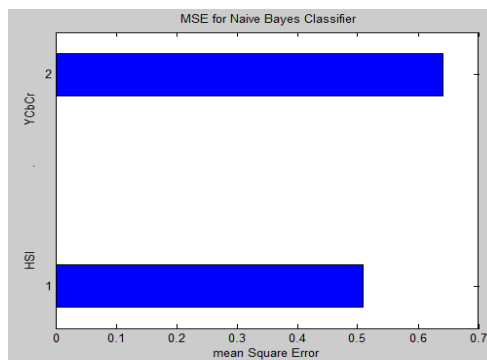

Figure 5. Graph showing Mean Square Error in case of HSI and YCbCr color space

The graph above shows that the mean square error obtained in case of HSI color space is 0.5590 and in case of YCbCr is 0.6210. The results show that HSI model best classifies the data set using naïve bayes classifier as compared to YCbCr model.

VII.   CONCLUSION AND FUTURE SCOPE

G. Conclusion:
The images of the cotton were collected from the cotton mill. The contaminants present in the cotton deteriorates the quality of cotton fiber so it s necessary to detect the contaminants from the cotton and then classify the contaminants. We used image processing tools to detect and classify different contaminants from the cotton fiber.

The research presents the implementation and comparative analysis of the HSI and the YCbCr color spaces for the detection and classification of contaminants and the foreign fibers from the cotton. Graph shows the comparison between the

HSI and the YCbCr color space clearly on the basis of time taken for detecting and classifying contaminants.Various experiments has been carried out on different images of cotton having different contaminants like bark , leaf, hair, nylon. etc. the time taken in whole algorithm in HSI color space is also less as compared to Ycbcr color space as shown in graphs above.

For classification, the texture features of images are extracted and naive bayes classification algorithm used to classify different classes of contaminants like bark, leaf, hair and nylon. It was found that training the model with a large number of test data and with fast training algorithm would greatly enhance the accuracy and hence the reliability of the system.The study has shown that the error rate of this classifier is near to zero.

H. Future Scope
- We can explore more algorithms and techniques for the feature extraction and classification of cotton contaminants to further improve the accuracy of the identification system.
- We can further improve the system by reducing the complexity. The main objective could be to find the best algorithms which optimize the performance and complexity.
- The accuracy of classifier can also be enhanced by using more and equal number of training patterns.

REFERENCES

[1]. R N Narkhedkar S S Lavate, Cotton contamination removal systems in blow room,june 2011

[2]. Zhenwei Su, Gui Yun Tian, and Chunhua Gao, "A machine vision system for on-line removal of contaminantsin wool", *Mechatronics*, Elsevier, Kidlington Oxford, 2006, pp. 243-247.

[3]. Liwei Zhang, Abbas Dehghani, Zhenwei Su, et al, "Realtime automated visual inspection system for contaminant removal from wool", *Real-time Imaging*, Elsevier, St. Louis, 2005, pp. 257-269.

[4]. Boshra D. Farah, "Image processing in the measurement of trash contents and grades in cotton", *Proceedings IWISP '96*, Elsevier, Manchester, 1996, pp. 629-632.

[5]. Bidan Li, Tianhuai Ding, and Dongyao Jia, "Design of a sophisticated foreign fiber separator", *Transactions of the Chinese Society for Agricultural Machinery (in Chinese)*, Editorial department of Transactions of the Chinese Society for Agricultural Machinery, Beijing, 2006, pp. 107-110.

[6]. A. Pai, H. Sari-Sarraf, E.F. Hequet, "Recognition of cotton contamination via X-ray microtomographic image analysis," IEEE Transactions on Industry Applications, vol. 1, 2004, pp.77-85.

[7]. Mingxiao Ding, Wei Huang, Bing Li, Shaohong Wu, Zhiqiang Wei, Yunkuan Wang, **"**An Automated Cotton Contamination Detection System Based on Co-occurrence Matrix Contrast Information*",*2009 IEEE.

[8]. Chengliang Zhang, Xianying Feng, Lei Li, Yaqing Song ,**"**Detection of Foreign Fibers in Cotton on the Basis of Wavelet", 2nd International Conference on Signal Processing Systems (ICSPS),2010 IEEE.

[9]. Zhang, Xianying Feng, Lei Li, Yang Song**, "**Identification of Cotton Contaminants Using Neighborhood Gradient Based on YCbCr Color Space", 2010 2nd international conference on signal processing systems, IEEE.

[10]. Pooja Mehta, Naresh Kumar, "Detection Of Foreign Fibers And Cotton Contaminants By using Intensity And Hue Properties*",* International Journal of Advances in Electronics Engineering, 2010.

[11]. Chai, K.; H. T. Hn, H. L. Chieu; "Bayesian Online Classifiers for Text Classification and Filtering", Proceedings of the 25th annual international ACM SIGIR conference on Research and Development in Information Retrieval, August 2002, pp 97-104 .

Authors Profile

Aditi Sachar received her Bachelor's degree from Model Institute of Engineering & Technology (Jammu University) during the year 2009 and now she is pursuing Mtech in computer science and engineering from Swami Vivekananda Institute of Engineering and Technology (Punjab Technical University).

Sugandha Arora obtained her Master degree (M.Tech) in Computer Science and Engineering from Kurukshetra University Kurukshetra in 2007, India and Bachelor's degree in Information and Technology from Kurukshetra University Kurukshetra in 2005.She has worked as Lecturer in Amity University, Noida. She then joined Surya Group of Institutions, Rajpura as Assist Professor Computer Science and Engineering till Jan 15, 2010. Presently, She is working as Assistant Professor in Computer and Engineering department with Swami Vivekananda Institute Engineering and Technology.