

# **Feature Identification in Wooden Boards Using Color Image Segmentation**

Srikathyayani Srikanteswara

Thesis submitted to the faculty of the  
Virginia Polytechnic Institute and State University  
in partial fulfillment of the requirements of the degree of

Master of Science  
in  
Electrical and Computer Engineering

Richard W. Conners, Chairman

A. Lynn Abbott

D. Earl Kline

October 21, 1997

Blacksburg, Virginia

Keywords: Image Processing, Image Segmentation, Wood Applications, Defect Detection

Copyright 1997, Srikathyayani Srikanteswara

# **Feature Identification in Wooden Boards Using Color Image Segmentation**

Srikathyayani Srikanteswara

(ABSTRACT)

Many different types of features can appear on the surface of wooden boards, lineals or parts. Some of these features should not appear on the surfaces of wood products. These features then become undesirable or removable defects for those products. To manufacture these products boards are cut up in such a way that these undesirable defects will not appear in the final product. Studies have shown that manual cut up of boards does not produce the highest possible yield of final product from rough lumber. Because of this fact a good deal of research work has been done to develop automatic defect detection systems. Color images contain a lot of valuable information which can be used to locate and identify features in wood. This is evidenced by the fact that the human color vision system can accurately locate and identify these features. A very important part of any automatic defect detection system based wholly or in part on color imagery is the location of areas that might contain a wood feature, a feature that depending on the product being manufactured may or may not be a defect. This location process is called image segmentation. While a number of automatic defect detection systems have been proposed that employ color imagery, none of these systems use color imagery to do the segmentation. Rather these systems typically average the red, green, and blue color channels together to form a black and white image. The segmentation operation is then performed on the black and white image. The basic hypothesis of this research is that the use of full color imagery to locate defects will yield better segmentation results than can be obtained when only black and white imagery is used. To approach the color wood image segmentation problem, two conventional clustering procedures were selected for examination. Experiments that were performed clearly showed that these procedures, ones that are similar in flavor to other unsupervised clustering methods, are unsuitable for wood color image segmentation. Based on the experience that was gained in

examining the unsupervised clustering procedures, a model based approach is developed. This approach is based on the assumption that the distribution of colors in clear wood is Gaussian. Since boards that are used by the forest products secondary manufacturing industry are all such that most of their surface area is clear wood, the idea is to use the most frequently occurring colors, i.e., the ones that must represent the most likely colors of clear wood, to estimate the mean and covariance of the Normal density function specifying the possible colors of clear wood. Deviations from this model in the observed histogram are used to identify colors that must be caused by features other than clear wood that appear on the surface of the board.

## **Acknowledgments**

I would like to thank my advisor, Dr. Richard W. Conners for his support and encouragement throughout my research. I would like to express my deep appreciation to him for his support and ideas without which this work would not have been successful. I also thank my committee members, Dr. Kline, for educating me about various aspects in the wood industry and providing me with crucial information towards my thesis and Dr. Abbot for all his help and cooperation. I thank them both for being in my committee. I am grateful to Mr. Phillip Araman for providing me with financial assistance towards my research and also helping me with my research.

I would like to thank Paul LaCasse for helping me scan the images of the boards, Panos Arvanitis the system administrator, and all the other members of the SDA Lab, Dr. Thomas Drayer, Frank Qiu, Qiang Lu, Yuhua Cui, and William King. I also thank all my friends for being so understanding and helpful.

I would like to express my gratitude to my parents for their encouragement throughout my studies and for always motivating me to be the best. I dedicate this work to them.

## Table Of Contents

<b>CHAPTER 1. INTRODUCTION.....</b>	<b>1</b>
1.1 Need for Automatic Feature Identification .....	1
1.2 Existing Defect Detection Processes and their Deficiencies.....	2
1.3 Motivation .....	4
1.4 Objectives and Scope of the Work .....	5
<b>CHAPTER 2. BACKGROUND .....</b>	<b>8</b>
2.1 Steps Involved in Processing Lumber.....	8
2.2 Features on the Wood Surface.....	9
2.3 Brief Note on Existing Automatic Defect Detection Systems .....	10
2.4 Segmentation .....	14
2.4.1 <i>Color Image Segmentation of Images of Wooden Boards</i> .....	15
2.5 Clustering .....	16
2.5.1 <i>Applicability of Clustering for Segmentation of Wood Images</i> .....	16
2.5.2 <i>Tradeoffs in Clustering</i> .....	17
2.6 Selection of the Clustering Algorithms.....	18
2.7 Multispectral Clustering Technique.....	19
2.7.1 <i>Description of the Algorithm</i> .....	19
2.8 Filtering Techniques .....	23
2.8.1 <i>Averaging Filter</i> .....	24
2.8.2 <i>Gaussian Filter</i> .....	24
2.8.3 <i>Median Filter</i> .....	25
2.8.4 <i>Scale Space Filtering</i> .....	25
2.9 Conclusions .....	26
<b>CHAPTER 3. METHODS.....</b>	<b>27</b>
3.1 Description of the Stages of Preprocessing .....	27
3.1.1 <i>Generation of the Histogram</i> .....	27
3.1.2 <i>Generation of 2-Dimensional Histograms</i> .....	31
3.2 Clustering a Histogram to Obtain a Map .....	31
3.3 Obtaining the Segmented Image.....	32
3.4 General Steps Followed to Segment an Image Using Clustering Techniques on the Histogram .....	32
3.5 Brief Note on the Boards Used for Testing .....	33
3.5.1 <i>Southern Yellow Pine</i> .....	33
3.5.2 <i>Red Oak</i> .....	33
3.5.3 <i>Yellow Poplar</i> .....	34
3.5.4 <i>Description of Some of the Most Commonly Found Features and their Significance</i> .....	34
<b>CHAPTER 4. MULTISPECTRAL CLUSTERING TECHNIQUE.....</b>	<b>36</b>
4.1 Modified Algorithm.....	37
4.2 Effect of varying the Initial Threshold $T_i$ .....	38
4.3 Analysis of the Effect of the Initial Threshold $T_i$ .....	47
4.4 Multiple Iterations .....	49
4.5 Histogram Characteristics.....	53
4.5.1 <i>Noise in the Histograms</i> .....	57
4.6 Filtering the Histogram.....	65
4.7 Scale Space Filtering Approach to Clustering .....	66
4.8 Drawbacks of Filtering .....	67
4.9 Evaluation of the Algorithm .....	69
4.10 Conclusions.....	71
<b>CHAPTER 5. HISTOGRAM MODELING TECHNIQUE.....</b>	<b>72</b>
5.1 Characteristics of Clearwood Histograms.....	74
5.2 Developing a Model Histogram.....	77
5.3 Description of the Algorithm.....	79
5.3.1 <i>Algorithm in 2-Dimension when Majority of the Board is Clearwood</i> .....	79
5.3.2 <i>Extending the Algorithm</i> .....	87

5.4	Effect of the Various Parameters on the Segmentation.....	94
5.4.1	<i>Number of channels used: 2-Dimensional, 3-Dimensional Histograms.....</i>	94
5.4.2	<i>Percentage of the original histogram used to create the model: p.....</i>	94
5.4.3	<i>Selecting the region from where the peak of the histogram is obtained.....</i>	95
5.4.4	<i>Value of the threshold <math>Th</math> used on the Weighted histogram <math>W</math>.....</i>	96
5.4.5	<i>Value of Offset.....</i>	96
5.5	Conclusions.....	97
<b>CHAPTER 6. RESULTS AND EVALUATION OF THE HISTOGRAM MODELING TECHNIQUE.....</b>		<b>98</b>
6.1	Description of the experiments.....	98
6.2	Results of Segmentation.....	100
6.2.1	<i>Red Oak.....</i>	100
6.2.2	<i>Southern Yellow Pine.....</i>	101
6.2.3	<i>Yellow Poplar.....</i>	104
6.3	Comparison of the Model Based Approach with the B/W Segmentation Algorithm.....	109
6.4	Conclusions.....	114
<b>7</b>	<b>SUMMARY AND CONCLUSIONS.....</b>	<b>115</b>
7.1	Recommendations for Future Research.....	117
<b>REFERENCES.....</b>		<b>121</b>
<b>VITA.....</b>		<b>125</b>

## List of Figures

Figure 2.1: Processing of hardwood lumber	9
Figure 2.2: Stages in an automatic defect detection system	13
Figure 2.3: Demonstration of 8 connectivity in 2 dimensions	20
Figure 2.4: Multispectral clustering technique in 1 dimension	21
Figure 2.5: Simple Averaging Filter	24
Figure 2.6: Median Filtering	25
Figure 3.1: (a) Image of a pine board p7a.dat, (b) Right edge that was detected, (c) Detailed view of the edge	30
Figure 4.1: Demonstration of 4-connectivity in 2 dimensions	41
Figure 4.2: (a) Image of a Pine board p8.dat, (b - e) Effect of setting $T_i = \text{mean}, 100, 435, \text{ and } 600$	42
Figure 4.3: Separation of Clusters	44
Figure 4.4: (a) Image of Yellow Poplar board y3a.dat, (b - e) Effect of setting $T_i$ at mean, 500, 800, 1000	46
Figure 4.5: (a) Image of Oak board c104a.dat, (b) Segmentation result with $T_i = 500$	48
Figure 4.6: Image of a Yellow Poplar board y2a.dat	51
Figure 4.7: Result of segmentation of Figure 4.6 after 3 iterations	52
Figure 4.8: Oak Board c116b.dat	55
Figure 4.9: (a) Two dimensional r-b histogram of oak board in Figure 4.8, (b) Two dimensional r-g histogram of oak board in Figure 4.8	56
Figure 4.10: Concept of a line through a 2-dimensional histogram	57
Figure 4.11: (a) variation in the 3-dimensional histogram along a line in 3-dimensional space, (b) line in 3-dimensional Space	59
Figure 4.12: (a) Two dimensional r-b histogram of pine board p8.dat in Figure 4.2(a), (b) Contour plot of the r-b histogram (c) Two dimensional r-g histogram of pine board p8.dat in Figure 4.2(a), (d) Contour plot of the r-g histogram	61
Figure 4.13: (a) Two dimensional r-b histogram of the yellow poplar board y3a.dat of Figure 4.4(a), (b) Contour plot of the r-b histogram (c) Two dimensional r-g histogram of the yellow poplar board y3a.dat of Figure 4.4(a), (d) Contour plot of the r-g histogram	64
Figure 4.12(a): 3x3 filter (b): Demonstration of 2D filtering	68
Figure 5.1: Board with only clearwood regions	75
Figure 5.2 (a): b-g histogram of the board in Figure 5.1 5.2, (b): r-b histogram of the board in Figure 5.1	76
Figure 5.3: Schematic diagram of Figure 4.13a, (r-b) histogram of yellow poplar board y3a.dat	78
Figure 5.4: Schematic diagram of Figure 4.9a, (r-b) histogram of pine board p8.dat	78
Figure 5.5: (a) Original histogram of an Oak board c104a.dat (Figure 4.5 a), (b) Modeled histogram using 85% of original histogram (c) Region $R$ , 85% of original histogram (d) Weighted difference using $offset = 1$ (e) Final segmentation in r-b color space using $threshold = 60$ (f) Segmented image of Oak board in Figure 4.5 (a)	86
Figure 5.6: (a) Image of a Pine board p5.dat (b) Two dimensional b-g histogram of the pine board in Figure 5.6(a), (c) Contour plot of the b-g histogram (d) Two dimensional r-g histogram of the pine board in Figure 5.6 (a), (e) Contour plot of the r-g histogram (f) Segmented image of pine board p5.dat shown in Figure 5.6a	93
Figure 6.1: Segmented Images of pine board of Figure 4.2a using the model based approach	103
Figure 6.2: (a) Segmented image of yellow poplar board of Figure 4.4	106
Figure 6.3: (a) Original image of a yellow poplar board y1a.dat, (b) Segmented image with heartwood-sapwood separation, (c) Segmented image without heartwood-sapwood separation	107

Figure 6.4: Segmented image of yellow poplar board in Figure 4.6 (a)	108
Figure 6.5: Segmented image of Pine board in Figure 4.2 (a), using b/w segmentation	111
Figure 6.6: Segmented image of Yellow Poplar board in Figure 6.3 (a), using b/w segmentation	112
Figure 6.7: Segmented Image of Yellow Poplar board in Figure 4.6 (a) using b/w segmentation	113
Figure 7.1: Hypothetical map of the color space indicating the probable presence of various features	119

## List of Tables

Table 4.1 Effect of varying the threshold for p8.dat (Southern Pine)	40
Table 4.2: Effect of varying the threshold for y3a.dat (Yellow Poplar)	41
Table 6.1: Parameters for Red Oak	100
Table 6.2: Parameters for Southern Yellow Pine	102
Table 6.3: Parameters for Yellow Poplar with Heartwood-Sapwood Separation	105
Table 6.3: Parameters for Yellow Poplar without Heartwood-Sapwood Separation	105

# Chapter 1. Introduction

## 1.1 *Need for Automatic Feature Identification*

Wood is the primary material from which many high-demand products are made. It is used as a structural building material; as a finishing building material as in doors, flooring and windows; as a packaging material; and as a material for making finished products like household furniture and cabinets. All these varied products involve numerous steps in creating a finished product from hardwood lumber [CON97]. Each product, depending on the application, and quality of the product requires a different look on the surface of the wood. In some applications the surface is covered with dark stains or paint, and the irregularities on the surface become less significant. However, in products like furniture, cabinets, wall or door surfaces, the surface features of wood are generally visible in the final product. In such applications, apart from other factors, appearance of the wood plays an important role in determining the quality and cost of the product.

Over thirty different kinds of features found on the surface of wood have been recorded [NHLA94] [HAR92]. These features are undesirable in some products depending upon the application and quality of the product. For instance, having dark knots or stains on a cabinet or door surface could drastically reduce the value of the product. In such a situation these features are considered as defects on the boards which go to make the final product. For this reason the words *defect* and *feature* are used interchangeably at times. Industries try to see to it that these undesirable features do not occur in their final products. When a tree is felled, it undergoes a series of processing steps until it is made into the final product. During these stages of processing, the earlier the defective regions (unwanted features) are removed from the wood, the better it is. The ideal time to identify and isolate these defects would be right after the tree is cut [CON97], but with existing technologies it is not possible to do so efficiently and the defects are generally isolated after the logs are cut. If all the unwanted features are removed in the early stages of processing, the clear areas can be utilized most efficiently, and this reduces the cost of manufacturing the product.

Once a board is cut from a log, it is edged and trimmed and then graded [NHLA]. These boards are then used to produce defect free dimension parts that will be used to create components for a finished product. Identifying and isolating these unwanted features is very important in order to produce good quality products. Boards are generally examined before they are cut in order to eliminate or minimize the defect area, and maximize utilization of the clear areas. However, the definition of what features constitute a defect is not absolute. Features in wood that are termed as defects for one application may be acceptable in another. A survey showed that most of the wood industries did not have a system that defines a set of features on the board surface as defects for a particular application [HUB90]. Huber et. al. surveyed 46 companies across the US and described at length the procedures followed by the industries to define 'what is acceptable' in the absence of a formal written system. Most of them relied on subjective verbal information communicated to the production workers. In such a situation there is a heavy responsibility on the production workers who visually examine the boards for defects and attempt to position the saw cuts to eliminate unacceptable defects while maximizing the yield of usable parts. Further, the worker has to memorize the specifications for each product while sorting and scanning the boards for defects. All this has to be performed at real time speeds. Such tasks are typically performed better by an automated vision system. It was found that on the average these employees obtained just about 68% of what could be called perfect results [HUB85]. They concluded that an automatic defect detection system need not be perfect to improve upon the current standards.

## ***1.2 Existing Defect Detection Processes and their Deficiencies***

A substantial amount of research has been done at the Brooks Forest Products Research Center at Virginia Tech in developing vision systems for the automatic processing of lumber. The areas of research include identification and classification of defects and the development of optimizing cutting algorithms. Automatic vision systems have been developed for different stages of wood processing including automatic rough mill systems. Elsewhere, other working

systems have used optical scanning techniques to locate and identify surface features that affect the quality of parts that are used in the production of furniture, flooring and cabinets [KLINE97].

An Automated machine vision system typically consists of the following subsystems or stages:

- 1) Scanning the board to obtain an image.
- 2) Preprocessing the image, which includes correcting any intensity or shade irregularities in the image.
- 3) Segmentation, which includes marking the various features on the board, which depending on the application, may or may not be a defect.
- 4) Recognition, which includes identifying and labeling the various features in the segmented image.
- 5) Classification of defects, which includes defining what features in the labeled image constitute a defect.
- 6) Decision making, which could be an instruction to the sawing, grading or any other similar process.

Most of the existing systems however, have one common drawback: the segmentation stage uses only the black and white (b/w) image instead of the full color information provided by a color camera. The advantage of using only the b/w information is that the data size is small and the processing is quick, making it simple and easy to handle. But in doing this, a lot of valuable information provided by the color camera is discarded.

Had the results of the segmentation using b/w images been highly acceptable, ignoring the color information could be justified. However, existing systems work quite well in detecting features like, dark colored knots, bark, wave and other obvious defects that can be separated using the b/w information alone. They do in fact perform quite poorly in detecting color variations, stains (blue stains) and light colored knots.

### **1.3 Motivation**

The performance of the vision system as a whole depends on the performance of each of its subsystems which were described above. In order to improve the performance of the overall system, it is important to improve the performance of each of its components. Improving the results of the segmentation stage will mean that the recognition stage has better results which in turn means better results from the recognition stage, and so on. Thus improving the segmentation stage will have far reaching effects, especially since the present state of technology does not provide the best possible segmentation results. Further it is also reasonable to say that the color information that is being discarded could be one of the reasons why the segmentation results are not truly satisfactory.

In this thesis, the use of the color information in the segmentation stage, to identify defects is examined. There are several reasons for studying color image segmentation techniques. First, the nature of the wood surface is such that most of the defects have a different color from that of clear wood. Different color characteristics in different features means that it is possible to separate a defect from clear wood in 3 dimensional red, green and blue (r, g, b) color space even if they lie on the same point on the b/w line. This is especially true for some features like stain and discoloration that may be very hard to identify in the b/w image. Theoretically it should thus be possible to obtain better segmentation results using the color information.

Secondly, fine furniture generally requires a uniform color, especially if they are lightly stained. In some species like yellow poplar, there is a large color variation particularly between the heartwood and sapwood regions. While these features are not generally classified as defects, and do not affect the grade of the board, they could adversely affect the quality of the final product. It depends on the nature and extent of the color variations and also the application of the final product. If the color variations between different regions of the board are unacceptable, stains are used to mask the color variations. Otherwise, the board is cut up such that each part is uniform in color and the parts are eventually color sorted. If the color variations are very drastic, they are generally treated as paint grade material. It is important to identify these features in the board before they are cut, if uniformity in color is important. While b/w segmentation can detect

defects, it is not possible to use the color characteristics of the board to optimize the cutting process. Say, parts are cut from the board and they turn out to be defect free during the regular NHLA grading process. They could still contain color variations that could make the part unacceptable for certain applications as stated earlier.

Thirdly, automatic vision systems that rely wholly or partly on color information can be improved with the use of good color segmentation techniques. The performance of existing systems can be improved by using all the color information obtained and exploiting the full potential of the scanning subsystem. The disadvantage in trying to develop such a vision system is that, unlike b/w image segmentation, color segmentation has not been studied in depth and there is a limited amount of literature in this field. But the main disadvantage is that using a larger measurement space increases the complexity of the algorithm and consequently the computing power required. But in view of the major advances in the computer industry, this factor should soon cease to be a major deterrent.

#### **1.4 Objectives and Scope of the Work**

The main emphasis of the thesis is to gain a better understanding of the multidimensional color space in terms of the grading features found on the surface of wood. To address this emphasis area, the following objectives will be accomplished: 1) Explore the use of the color information and examine the use of conventional clustering techniques in segmenting the images of wood, and also to quantify the significance of this information in terms of improving the quality of segmentation. 2) Study and develop a new model based approach that can accurately segment regions in a multidimensional color histogram relating to some surface features on wood. 3) Evaluate the performance of this approach in terms of accuracy, complexity, and flexibility and compare it to an existing black and white segmentation algorithm used in a defect detection system.

Since the new model based approach must be useful in future real time systems, in addition to providing accurate results, the complexity of the algorithm has to be kept at a minimum.

Further, the segmentation algorithm should also have flexibility. In other words, different applications will have different specifications of what grading features are acceptable. It should be possible to adjust the algorithm to meet the needs of each specific application. The result of the algorithm will have to be a map of the board separating the surface into *defect* and *clearwood* regions. All those regions on the board that are not acceptable for a particular application (the application for which the algorithm has been tuned for) will have be marked as defects. This result can then be passed on to the next stage of a vision system, the recognition stage. The recognition stage can now use the results of segmentation to identify and classify the regions into the various defect categories.

If any feature identification has to be done based on color, the first place to start is with the full color histogram of the image. The full color histogram is a 3 dimensional histogram of the red, green and blue channels. The use of some of the preexisting techniques in segmentation will first be examined. In particular modifications of two conventional clustering algorithms will be investigated. The two techniques which will be studied are based on a multispectral clustering scheme proposed by Goldberg and Shlien [GOL78], and a scale - space filtering technique [WIT84]. Next, a model based approach will be developed after studying the use and applicability of the conventional segmentation (clustering) algorithms.

A model is developed for the histogram of the desired regions on the board surface, and the observed histogram is compared to this model. A difference measurement is developed that is used to determine the areas in the observed histogram that do not belong to the clear wood sections. This approach is based on the assumption that the distribution of colors in clear wood is Gaussian. Since boards that are used by the forest products secondary manufacturing industry are generally such that most of their surface area is clear wood, the idea is to use the most frequently occurring colors, i.e., the ones that must represent the most likely colors of clear wood, to estimate the mean and covariance of the Normal density function specifying the possible colors of clear wood. Deviations from this model in the observed histogram are used to identify colors that must be caused by features other than clear wood that appear on the surface of the board.

If acceptable segmentation can be obtained with smaller amounts of information (and complexity), it would be a definite advantage. Hence, the results of the algorithm will be examined using only two channels of information. The model based approach also provides flexibility in defining clear wood and defect areas. For instance, if a particular application is not affected by color variations (like heartwood and sapwood in yellow poplar), the parameters of the algorithm can be set such that they are not detected. On the other hand, the parameters can be adjusted such that these features are detected if they are critical to the application. The performance of an existing b/w segmentation algorithm will also be compared to the model based approach.

Automatic defect detection systems for wood have been developed and implemented using various technologies, but the use of color information in the segmentation procedure has often been overlooked. In particular, the use of the clustering techniques on the full color histograms has not been studied. The study of the applicability of the conventional techniques, and the other concepts and ideas developed in this thesis, will lead to a better understanding of the color characteristics of wood. The model based approach, along with the encouraging results obtained, should motivate research to exploit the full potential of the color information in existing defect detection systems in the wood industry.

## Chapter 2. Background

### 2.1 Steps Involved in Processing Lumber

Forest products manufacturing can be broken down into two basic processing steps. The first step, primary processing involves turning logs into a product that can be used by others. Primary processing include sawmillers that turn logs into lumber and veneer. The second processing step, secondary processing or secondary manufacturing, involves turning the products created by primary manufacturers into products that are sold to the buying public, i.e., doors, windows, cabinets, furniture, flooring and other household fixtures.

The typical components of a hardwood lumber sawmill include a debarker, a headrig, an edger, a resaw and a trimmer shown in Figure 2.1[KLINE92]. Though actual mill layouts can vary substantially, in general, logs enter a debarker to remove bark, and debris that may prematurely dull saw blades. Debarked logs wait on a log check for primary breakdown at the headrig. The headrig which employs a circular or a band saw, breaks the log down into boards. Boards with wane are routed to the edger where wane and other edge defects that effect the grade are removed. Boards from the headrig, edger, and gang resaw are then routed to the trim saw to remove end defects that affect grade. In certain cases, some gang resawn boards are routed to the edger to remove wane. In any case, nearly every board is processed at the trim saw. Finally edged and trimmed lumber is graded, sorted and stacked.

These boards are then used to produce defect free dimension parts that will be used to create components for a finished product. Identifying and isolating any unwanted features on the board is very important for producing good quality products. Regions on the board surface that are free from defects are called *clearwood*. Boards are generally examined before they are cut in order to eliminate or minimize the defect area, and maximize utilization of the clearwood areas.

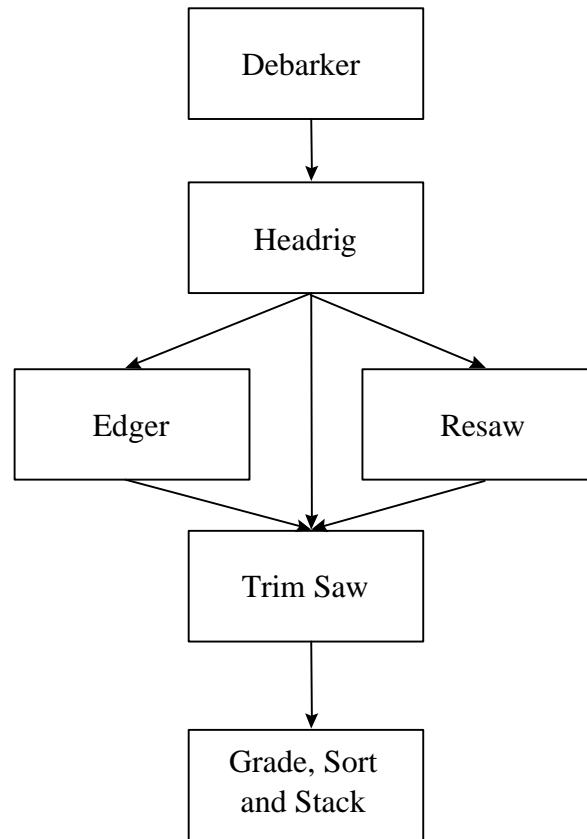


Figure 2.1: Processing of hardwood lumber

## 2.2 Features on the Wood Surface

There are a large number of features that are found on the surface of the boards [NHLA94]. Some of the features like decay, spike knots, unsound knots etc. affect the strength of the wood, and are termed as structural defects. Other features like color variation, stains, sound knots, pith etc., affect the grade of the board but do not weaken the structure of the material. There are yet other features like color variations on the board do not technically affect the grade of the lumber. But fine furniture and similar applications generally require a uniform color, which means that variations in color have to be considered when the board is cut.

The different types of defects and the relative frequency of their occurrences have been described in detail for different species of wood [WDB95, HAR92]. If non-structural defects are

present, the board can be used as *paint grade* material, and the defects can be concealed with the use of paints and dyes. Whereas if structural defects are present, they cannot be used even as paint grade material and the defects have to be eliminated when the board is cut. Each product, depending upon the application, and quality of the product requires a different look on the surface of the wood. This means, that features on the wood surface that are classified as defects for one application may be acceptable in another. For instance in applications where the surface is covered with dark stains or paint, the irregularities on the surface become less significant. In this case, features like color variations on the board will not be classified as defects, unlike if the final product was fine furniture, cabinets, wall or door surfaces. The definition of a *defect* is thus relative. A *defect* for a particular product can be defined as a subset of those *features* on the wood surface that are unacceptable in the final product. So when boards are examined before they are cut, defects for a particular product are eliminated. The board is cut up such that the defects on the board do not appear in the cup up sections that will be used to form the final product. Alternatively, if the board has a lot of unacceptable features, it will be used with stains and dyes to mask the defects. In cases where the features cannot be masked by stains, the board is classified as paint-grade material.

### **2.3 Brief Note on Existing Automatic Defect Detection Systems**

A substantial amount of research has been done at the Brooks Forest Products Research Center at Virginia Polytechnic Institute and State University, Blacksburg, in developing vision systems for automatic processing of lumber [KLN92, KLN97]. The areas of research include identification and classification of defects and using these results in optimizing cutting algorithms. Automatic vision systems have been developed in different stages of wood processing including automatic rough mill systems.

Elsewhere, other working systems have used optical scanning techniques to locate and identify surface features that affect the quality of parts that are used in the production of furniture, flooring and cabinets [KLN97]. Optical scanning methods include the use of black/white cameras, color cameras, and spectrometers to measure the intensity and color of the

reflected light [CON97]. These methods are capable of detecting surface features such as knots, splits, wane, stain color, and grain pattern [GUD95]. Feature recognition employing optical sensing techniques have been shown to be readily automated [CON92, KLI93]. The main disadvantage of optical-based machine vision systems is that they are extremely sensitive to wood species and also condition of the wood like moisture content, roughness etc.

Using the spectral characteristics of wood to identify features on the surface has been examined by Lebow et. al. [LEB96, MAR94]. They develop a database of spectral reflectance curves and model and classify the spectral reflectance curves by feature type. This technique however, has not led to a commercial vision system so far.

Many surface features in wood are directly related to changes in wood density [CON97]. Such features include knots, splits, decay, holes and abrupt changes in slope of wood grain. Machine vision systems have been developed based on density scanning techniques using ultrasound, mcrowaves, nuclear magnetic resonance (NMR), and x-ray [SZY81, POR92]. Scanning systems that measure wood density overcome some of the deficiencies of optical scanning systems. However, they cannot detect features relating to color variations in wood such as stain, heartwood, sapwood etc.

Automated machine vision systems that use some form of scanning technology typically consists of the following subsystems or stages [CON97] as shown in Figure 2.2:

1. Scanning the board: The board is scanned to obtain a visual image. The scanning process could be done using a color or a black/white camera.
2. Preprocessing the image: This includes correcting any intensity or shade irregularities in the image. This corrects the inherent problems associated with the scanning technology used, making the image a true replica of the actual features of the board.
3. Segmentation: This includes marking the various features on the board, which depending on the application, may or may not be a defect. At the end of this stage, the clearwood areas are separated from the defect areas on the board.

4. Recognition: This step involves identifying and labeling the various features in the segmented image. For instance, the various regions marked by the segmentation process are classified under a particular category like, knot, wave, pith, etc.
5. Classification of defects: This process defines what features in the labeled image constitute a defect. For each application there might be a set of features that are unacceptable. Those defects alone are identified on the labeled image. At the end of this stage, all the unwanted areas on the board are marked.
6. Decision making: This is generally an instruction to a mechanical device based on the final defect map of the board. For instance it could be an instruction to the sawing, grading or any other similar process. In cases where the board is entirely unsuitable, it could be classified under a different category for use with paints and dyes, or could be discarded.

In most of the existing commercial systems, the segmentation stage uses only the black/white image instead of the full color information provided by a color camera. The advantage of using only the black/white information is that the data size is small and the processing is quick, making it simple and easy to handle. But in doing this, a lot of valuable information provided by the color camera is being discarded. However, representing and interpreting the color information lead to much more complex algorithms

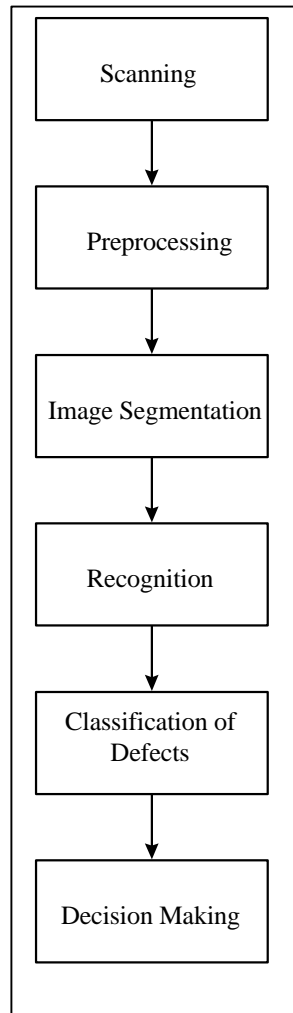


Figure 2.2: Stages in an automatic defect detection system

Had the results of the segmentation using black/white images been highly acceptable, ignoring the color information could be justified. However, existing systems work quite well only in detecting features like, dark colored knots, bark, wane and other obvious defects that can be separated using the black/white information alone. They do in fact perform quite poorly in detecting color variations, stains (blue stains) and light colored knots. To overcome the shortcomings of the black/white segmentation process, multiple scanning technologies have been used in defect detection systems [KLN97, CON97] that use X-ray, laser and b/w images. The

information obtained from all the three sources are combined to obtain the final defect map of the board.

## **2.4 Segmentation**

Segmentation or feature detection is a process of subdividing an image into its constituent parts or objects [GON92]. The constituent parts or regions, should correspond to structural units in the image, or distinguish objects of interest [RUS92]. A region is defined as an area with homogeneous spatial properties [YNG86]. The level to which this subdivision is carried depends on the problem being solved. That is, segmentation should stop when the objects or regions of interest in an application have been isolated. Segmentation is a form of pixel classification, i.e., each pixel is assigned to a class of pixels based on some property of the pixel. As a result of segmentation, the regions should be homogeneous with respect to the segmentation criterion such as uniformity in gray level intensity or texture [HSN91].

There have been numerous classifications of segmentation [HRL85, GON92], but they can be broadly divided into two categories based on the approach: (1) edge based and (2) region based [YNG86]. In edge based methods, the local discontinuities are detected first and then connected to form longer, hopefully complete, boundaries. In region based methods, areas in the image with homogeneous properties are found first. These homogeneous areas are then used to define the segmentation boundaries. The edge-based methods are based upon recognizing properties of dissimilarities in an image, whereas region based methods utilize similarity properties [TRV86]. The edge based methods include edge and line detection techniques. Some of the commonly used edge detection techniques are gradient method, surface fitting, template matching, second derivative methods and thresholding and thinning [YNG86]. The line detection methods include grouping by using Hough transforms and other simple line linking techniques [GON92]. Region based methods include thresholding and recursive segmentation, region growing, split and merge methods and clustering [GON92].

### **2.4.1 Color Image Segmentation of Images of Wooden Boards**

A number of color image segmentation techniques have been proposed using region based methods as well as edge based methods [GON92, SAR93], [IVI95], [HRL85], [FUK80], [YGN86]. These algorithms are designed for images of scenery or objects etc. However, images of wood do not fall under any of the categories of images examined by the previous researchers. Hence, the problem of feature identification in images of wood differs from these routine images. While the general principles of segmentation can be applied to the images of wood, the specific techniques that have been developed are neither optimum, nor are they successful in segmenting images of wood.

There are a wide range of colors, features, shapes and objects in a general image, while the images of wood, are not as diverse as the general real world images. However, the surface of wood is not homogeneous and is characterized by a high frequency grain pattern and a low frequency color variation that complicates the segmentation process. In spite of this, the color characteristics of wood can be exploited in the segmentation process, making it more accurate, rather than employing a generalized segmentation algorithm. In some respects the process can be simplified greatly based on the nature of the problem. There is yet another problem with most conventional image segmentation techniques, and in particular with color image segmentation techniques. The purpose of developing a segmentation technique is to use it for automatic feature identification. This means that all the algorithms used for automatic feature identification will have to perform in real time, and most of the conventional schemes described above are not suitable for this purpose.

Color alone forms a major part in separating out the various features of wood, unlike in images of objects and scenery. This is why clustering techniques will be examined in greater detail. Brunner et. al. [BRU90] have investigated the effect of using different color coordinate systems in representing the surfaces of wooden boards and how they impact the defect detection process. They conclude that no one system can be suitable for all applications and lighting geometry and source, camera system, and color space have to be carefully chosen.

## **2.5 Clustering**

Clustering is one of the region based segmentation techniques. Cluster analysis is a tool of exploratory data analysis that attempts to assess the interaction among patterns by organizing the patterns into groups or clusters such that patterns within a cluster are more similar to each other than are patterns belonging to different clusters. Many definitions have been proposed to define an ideal cluster, but no single definition is adequate [YNG86]. A cluster is said to be a set of entities that are *alike*, and entities from different clusters that are not alike. Clusters may also be described as connected regions of a p-dimensional space containing a relatively high density of points, separated from other regions by a region containing a relatively low density of points. It is clear that the user's prior conception determines what a cluster means and sets the goals for a clustering method [YNG86].

Most papers on clustering present experimental evidence of results to illustrate the effectiveness of their clustering processes. But to date, no objective quantitative measure of clustering performance is available, although a lot of effort has been expended toward that end. In clustering applications, generally the modes are first located, that is the local maximum of the probability density is obtained if the number of classes is known [BOW92]. When the number of classes is unknown, unsupervised learning methods will have to be used. Usually, an estimate of the number and location of modes is obtained, that is the natural grouping of patterns is found. All defect detection system for wood, will have to use unsupervised learning methods since the number of (types of) defects that could be present on the boards is unknown.

### **2.5.1 Applicability of Clustering for Segmentation of Wood Images**

For images of wood, the objective is to cluster data obtained from the image into different regions like, clearwood and defect clusters, so as to obtain a meaningful segmentation of the original image. Spatial clustering refers to grouping the individual pixels (clusters) of an image into regions. However, clustering, if performed on the histogram of the image rather than the image itself has a tremendous advantage over other forms of segmentation. The size of the measurement vector remains constant irrespective of the size of the image. This feature is

particularly useful if the size of the image is quite large which is true in most applications in the wood industry. For instance the average size of the images used was around 5000x900 pixels, while the size of the full color histogram, using 6 bits to represent each channel is 262144. This leads to a reduction in data size of around 17 times.

It is true that clustering cannot be used to segment all kinds of images, especially clustering of the histogram to perform segmentation on the image. There are however two reasons that make it possible to use this technique on the images of wood. First, the very nature of most of the defects are color based. This means that it should be possible to identify regions in the histogram that are a part of a defect and regions that are a part of clear wood. This implies that it should be possible to *cluster* the three dimensional color space corresponding to different regions on the wood surface. Second, most non-uniformities due to lighting can be removed in the pre-processing stage, after the image is obtained, using a shading correction algorithm [SAW77]. It is thus possible get colors that are almost a true representation of the wood surface.

There is one disadvantage in using clustering on the histograms to segment the images. It is possible that a particular color can be seen in a defect as well as in the grain pattern and it is not easy to differentiate these grading features from the grain pattern. However, such errors should not greatly undermine the quality of segmentation since the features on wood are essentially color based. Further no segmentation method guarantees perfect results. It is thus worthwhile examining this technique considering the reduction in complexity when compared to other edge based or region based methods.

### **2.5.2 Tradeoffs in Clustering**

It was mentioned earlier that it is meaningful to use clustering because of the nature of the wood surface. This is based on the assumption that defects can be separated on the basis of color alone. However, wood has a heterogeneous surface. It is characterized by a low frequency variation in color along the length and width of the board and a high frequency grain pattern. These are an integral part of the wood surface and do not constitute part of the defects. This makes the clustering process more complicated. Consequently it may not be possible to obtain

perfect separation of defects in all boards. There could be situations where a color that is very prominent in a defect, is also seen in small quantities in other areas of clearwood. But if the color is very prominent in a defect, it has to be marked as a defect color. This means that some clearwood regions can be included as defects. Instead, if we chose to mark this color as a 'clearwood' color, we are ignoring the entire defect area, which could be counter productive. If the area of the defect is large, it might be more useful to mark this color as a defect, and lose some clearwood, rather than miss part of the defect. On the other hand if the defect is insignificant and does not greatly undermine the quality of the board, it would be wise to classify this color as a clearwood color. The whole problem of clustering is to decide where to set this optimum threshold that minimizes these erroneous classifications and achieves the best trade off.

## **2.6 Selection of the Clustering Algorithms**

A wide variety of clustering algorithms exist and some of them are for multispectral data too. [GOL76, FUK80, WIL90]. Haralick et. al. [HRL85] give a very good summary of the existing techniques. The use of these existing techniques are first examined since, the conventional application of these techniques have never been applied to histograms of images of wood. To use these algorithms in the area of wood processing, it has to meet certain criteria. First, since the total number of clusters are unknown, any supervised learning methods cannot be used. Secondly, the algorithms cannot be slow since they have to be used in a real time defect detection system. The algorithms, at least, have to meet the minimum speed requirements of a real time vision system. Thirdly, the algorithms should also have some amount of flexibility in them so that the defect detection system can be tailored to meet the needs of a specific wood industry application.

Based on an initial survey of the existing clustering algorithms, two of them were identified to be investigated in detail. These algorithms were identified after a survey of the existing techniques and the results that the authors claimed to have obtained. The first is a multispectral clustering algorithm that is an iterative split and merge process [GOL78]. The second approach is based on scale space filtering [WIT84] described in Section 2.8.4, under filtering techniques.

The concepts used in these two approaches form the basis for most of the conventional clustering techniques. Studying and understanding these techniques will provide a good understanding of how the conventional techniques will perform when applied to the histograms obtained from images of wood.

## **2.7 Multispectral Clustering Technique**

The multispectral clustering technique was proposed by Goldberg and Shlien for four dimensional LANDSAT imagery[GOL78]. There are several reasons why this approach was selected to be examined in detail. First the clustering algorithm is applied to multidimensional measurement vectors of dimensionality greater than three. Clustering algorithms proposed for segmenting images generally operate in the spatial domain. This means that the data set has just two dimensions [FUK80, HRL85]. Wilson et. al. [WIL90] proposed a clustering scheme for segmenting images, using a clustering technique. They claimed that the algorithm can be extended to multidimensional data, but the images they used for testing were all b/w images. Clustering schemes proposed for higher dimensions are largely for analyzing data and not necessarily images [EIG74, GAT89].

Secondly, there are no assumptions made on the number of clusters. The potential of the multispectral clustering algorithm also comes from the fact that Goldberg et. al. have proposed a solution to the elusive problem of obtaining the optimum number of clusters in a measurement vector. Thirdly, they have a certain amount of flexibility in their algorithm. This is an interactive method based on splitting and merging the clusters formed in the histogram, until the results obtained are to the user's satisfaction.

### **2.7.1 Description of the Algorithm**

The observed histogram is assumed to be a mixture of several multidimensional gaussian distributions with unknown mean vectors, covariance matrices, and a-priori probabilities. Estimates of the gaussian parameters are determined ignoring the overlap of the neighboring distributions. The theoretical histogram is then calculated by integrating numerically, the

probability density function in each of the cells of the histogram, and the likelihood ratio test is applied to measure the departure of the model from the observed data. A statistical measure, taking into account the number of degrees of freedom is defined and used to choose between alternate models. The actual steps followed by the algorithm will now be described.

Let  $\hat{H}(i, j, k)$  be the 3 dimensional histogram generated from the r, g, b color image.

1. Find the average (mean) number of pixels found in each color cell. Label this value as threshold  $T_i$ , i.e.,  $T_i = \text{Mean}$ .

$$\text{Mean} = \frac{\sum_{i=0}^n \sum_{j=0}^n \sum_{k=0}^n H(i, j, k)}{n^3}, \quad n = 2^m$$

where,  $m =$  number of bits used for each channel in the histogram. In this case  $m = 6$  and  $n = 64$ .

2. Threshold the histogram  $\hat{H}$ , and divide the cells of the histogram into two sets,  $S1$  and  $S2$  such that the cells in  $S1$  have values greater than or equal to  $T_i$ . The cells in  $S2$  have values lesser than  $T_i$ .
3. Find the connected components of  $S1$ , to give  $K$  clusters. Connected components are found by extending the concept of 8 connectivity to higher dimensions. 8 connectivity in two dimensions is demonstrated in Figure 2.3. As the name suggests, a cell will have 8 neighbors. In three dimensions a cell will have 26 neighbors, with the cell being in the center of a  $3 \times 3 \times 3$  cube.
4. Assign the cells of  $S2$  to the nearest cluster among the  $K$  clusters. Now the histogram is divided into  $K$  clusters. Steps 2 – 4 are illustrated for a one dimensional data set in Figures 2.4.

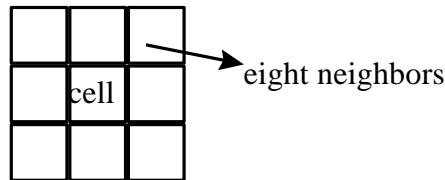


Figure 2.3: Demonstration of 8 connectivity in 2 dimensions

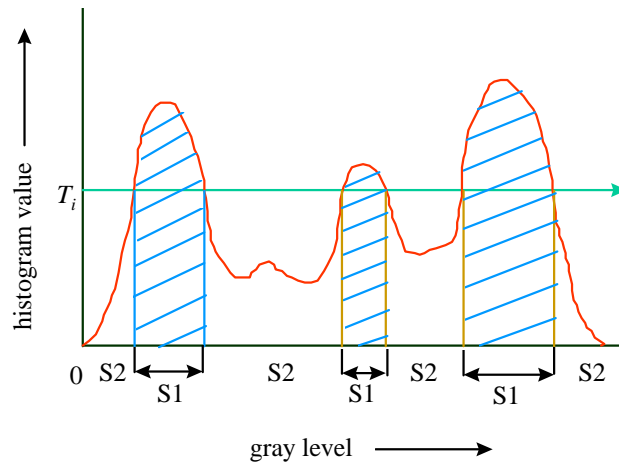


Figure 2.4a: Thresholding a histogram to obtain  $S1$  and  $S2$

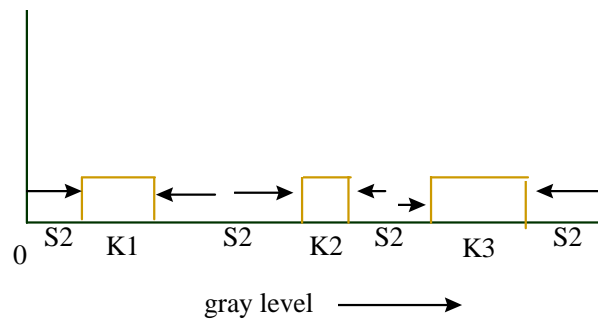


Figure 2.4b: Assign cells in  $S2$  to the nearest cluster in  $S1$  to form  $K$  clusters

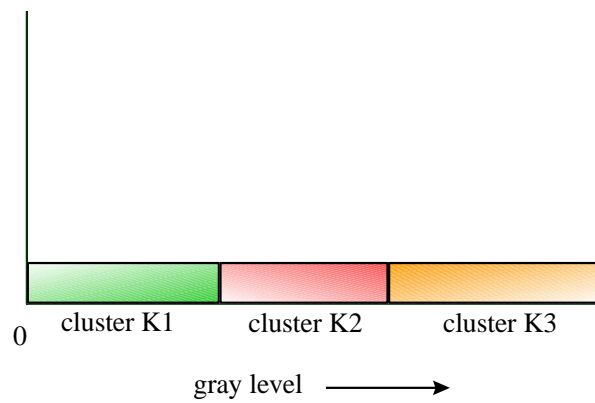


Figure 2.4c: Color space divided into  $K$  clusters

5. Fit a normal distribution into each of the  $K$  clusters. The  $K$  clusters are assumed to have unequal mean vectors  $\bar{m}_k$  and unequal covariance matrices  $\Sigma_k$  ( $k = 1, 2, \dots, K$ ) [GOL78]. The model is now

$$P(x) = \sum_{k=1}^K P(W_k) G(x | \bar{m}_k, \Sigma_k)$$

where,

$$G(x | \bar{m}_k, \Sigma_k) = \frac{1}{(2\pi)^p |\Sigma_k|^{1/2}} \exp[-\frac{1}{2} (x - \bar{m}_k)^t \Sigma_k^{-1} (x - \bar{m}_k)]$$

and  $P(W_k)$ ,  $k = 1, 2, \dots, K$  are the a priori probabilities. The mean vectors and the covariance matrices are evaluated while assuming the clusters do not overlap, i.e.,

$$\bar{m}_k = \frac{1}{N_k} \sum_{i=1}^{N_k} x_i^{(k)}, \quad \Sigma_k = \frac{1}{N_k - 1} \sum_{i=1}^{N_k} (x_i^{(k)} - \bar{m}_k)(x_i^{(k)} - \bar{m}_k)^T$$

$$P(W_k) = N_k/N \quad \text{and} \quad N = \sum_{k=1}^K N_k$$

where,  $x_1^{(k)}, x_2^{(k)}, \dots, x_{N_k}^{(k)}$  are the data vectors belonging to class  $k$ .

6. The likelihood ratio test is formulated by specifying the null and alternate hypotheses. In this case the null hypothesis assumes the samples originate from the theoretical distribution  $H(I)$  where,

$$H(I) = \int \int \dots \int_{I_l} P(x) d^n x,$$

$I_l$  is a single histogram element, the integration is performed over all the dimensions, and

$$P(x) = \sum_{k=1}^K P(W_k) G(x | \bar{m}_k, \Sigma_k).$$

The alternate hypothesis is here unspecified and consists of all other possible distributions.

The logarithm of the likelihood ratio for the two hypotheses is therefore given by

$$\Lambda = \sum_{l=1}^f \hat{H}(I_l) \ln \left( H(I_l) / \hat{H}(I_l) \right)$$

where, the summation is carried over  $f$  intervals  $I_l$  in which the observed histogram  $\hat{H}(I_l)$  is non-zero. If there are several unknown parameters in the model, then they can be chosen so as to maximize  $\Lambda$ . It has been shown that [GOL78] when the null hypothesis is true,  $-2\Lambda$  asymptotically approaches the  $\chi^2$  distribution with  $F$  degrees of freedom.

$$f = F + e + 1$$

where,  $F$  is the number of degrees of freedom, and  $e$  is the number of estimated parameters. The likelihood ratio test is used to determine how far a given model departs from the observed data. The  $p$  value is defined as the probability of the likelihood ratio statistic  $\Lambda$  being smaller than the observed  $\Lambda$  given that the null hypothesis is true. It is also the lowest significance level at which the null hypothesis can just be rejected.

7. Goldberg et. al. Claim that the best model is obtained when the  $p$  value has the lowest possible value. Clusters are split or merged interactively and the entire process is repeated until a satisfactory model is obtained.

## 2.8 Filtering Techniques

The multispectral clustering technique is susceptible to noise in the data set that could lead to false peaks. Thus the next approach to be examined involves filtering the data set to overcome the problems associated with noise. The second approach – the scale space filtering approach to clustering is also described in this section. However, before getting into the details of the scale space approach to clustering, the effects of filtering in general will be studied.

Several types of low pass filters exist which smooth the irregularities in a data set [GON92, YNG86]. Filtering techniques are extensively used in 1 and 2 dimensions. The commonly used two dimensional filters are discussed in the following sections. Extending these techniques to higher dimensions will be discussed in Chapter 4.

In two dimensions, the filter is represented as a two dimensional mask and is convolved with the data set to obtain the filtered data. If  $D$  is a two dimensional data set and  $F$  is a 3x3 filter, then the filtered data  $D_f$  at coordinates  $(a, b)$  is given by

$$D_f(a,b) = \sum_{i=0}^2 \sum_{j=0}^2 \sum_{k=0}^2 F(i,j) \times D(a-1+i, b-1+j)$$

### 2.8.1 Averaging Filter

One of the simplest forms of a low pass filter is the averaging filter, which is represented in 2 dimensions as (Figure 2.5)

1	1	1
1	1	1
1	1	1

Figure 2.5: Simple Averaging Filter

This filter merely averages the histogram over the neighborhood, giving equal weights to all the cells. There exists other filters which perform the same averaging operation, but with different weights for different cells. One such filter is the Gaussian filter, which is described in the next section.

### 2.8.2 Gaussian Filter

The Gaussian distribution, when used as a filter has some nice smoothing properties and serves as a very good averaging filter[GON92]. While the simple averaging filter gives equal weights to all the surrounding pixels, the gaussian filter uses the Gaussian distribution to select the individual weights for the mask. A multivariate normal (gaussian) distribution is given by

$$N(\bar{x}) = \frac{1}{(2\pi)^{d/2} |\Sigma|^{1/2}} \exp[-\frac{1}{2}(\bar{x} - \bar{m})' \Sigma^{-1}(\bar{x} - \bar{m})],$$

where,

$\bar{x}$  is a  $d$  - dimensional column vector

$\bar{m}$  is a  $d$  - dimensional mean-vector

$\Sigma$  is the  $d \times d$  covariance matrix

Also, in a given  $d$ -dimensional Gaussian distribution with  $n$  samples

$$\bar{m} = \frac{1}{n} \sum_{k=1}^n x_k$$

$$S = \frac{1}{n} \sum_{k=1}^n (x_k - \bar{m})(x_k - \bar{m})^t$$

The Gaussian filter in two dimensions is generated by a 2-dimensional Gaussian distribution with a mean at the center of the filter and variance defined by the covariance matrix. The extent of filtering is controlled by the covariance matrix.

### 2.8.3 Median Filter

A median filter is generally used to eliminate salt and pepper noise in the signal [GON92]. The value at the center of the mask is replaced by the median value of all the cells enclosed by the mask. For instance if a1 - a9 (Figure 2.6) represent a section of a 2 dimensional signal, a 3x3 median filter would replace the value at a5 by the median value of a1 - a9.

a1	a2	a3
a4	a5	a6
a7	a8	a9

Figure 2.6: Median Filtering

### 2.8.4 Scale Space Filtering

The concept of scale-space filtering [WIL90] is to repeatedly filter the signal to smooth out spurious peaks, while using the original signal to obtain the accurate location of the major peaks. The reason for doing this is that, the location of the peaks are shifted when the signal is passed through a filter. Roberts et. al. [STP95] estimates the optimum number of clusters by varying the scale over a broad range and estimating the number of clusters in each stage. A stable number of clusters over a range of variation in scale is considered as optimum clustering.

## **2.9 Conclusions**

Various segmentation techniques are present in the literature, which are either edge based or region based, and some of these techniques have been applied to color images as well. Clustering is a region based technique, and if performed on the histogram of the image rather than the image itself has a tremendous advantage over other forms of segmentation. The use of the existing clustering techniques needs to be examined first since, these techniques have never been applied to images of wood. A study is needed to investigate the possibility of segmenting color wood images by clustering the histogram. If it is possible, then the applicability of conventional clustering techniques needs to be examined, before a specific algorithm can be developed for wood. These questions will be addressed in Chapter 4.

The results of the studies performed in Chapter 4, suggested the need for an entirely new approach to the problem of segmenting wood images using clustering techniques on the color histograms. A novel histogram modeling technique is developed and discussed in Chapter 5, which gives very encouraging results.