

## Chapter 5. Histogram Modeling Technique

The conventional algorithms studied in the previous chapter did not perform very well, due to the nature of the problem and the size of the data. The size of the data made it impossible to use iterative algorithms in real time implementation. Most traditional clustering algorithms have the underlying assumption cluster centers correspond to significant peaks in the histogram. If this is not true then the only recourse is to use conventional iterative techniques in the hopes that such methods will improve the segmentation. The histograms of wood used in the study did have peaks and valleys, but clustering these histograms based on the above assumption does not lead to the desired segmentation. The histograms of wood have lots of peaks in the histogram, peaks that do not necessarily represent separate clusters. On the other hand, distinct peaks do not always represent separate clusters. This effect is compounded by noise in the image data (see Section 4.4.1). In addition to the presence of strong peaks that do not correspond to distinct clusters, a large number of defect clusters (in typical boards) are weak, especially when they are situated away from the main cluster. A defect cluster appears as a weak peak if the defect area (or undesirable areas) on the board is not very large, which is generally the case. In such a situation, the defect cluster will not have a distinct peak, but appears to be merged with the clearwood cluster. Only if there are distinct color variations in the board, as is the case in most Yellow Poplar boards, each feature has one or more distinct peaks.

Eliminating or reducing the effects of noise is not possible through filtering. It has the basic drawback of distorting data contained in the histogram, e.g., cluster centers, and removing small defect clusters entirely. The size of the histograms limits the number of iterations that can be used, since the segmentation must be performed in real-time. On smaller histograms, the segmentation results could probably be refined over a number of iterations even if conventional techniques are used. However, this is only a matter of conjecture since no experiments were performed to substantiate this conclusion.

Based on these findings, it is reasonable to say that the segmentation algorithm has to possess the following characteristics. First, it is desirable that the method be robust so that it will work on histograms generated from noisy image data. Second, the method should not be

iterative considering the size of the histograms. Both these factors indicate the need for a more basic and direct approach to the problem.

A segmentation technique is developed that can adapt itself to the nature of the histograms that are obtained from images of the wooden boards. This technique is based on developing a *model* or an ideal histogram for each board and measuring the difference between the modeled histogram and the actual histogram. The ideal histogram is the histogram that would be created from a board that is composed entirely of clearwood. The ideal histogram is modeled as a Gaussian or normal density function whose defining parameters are the mean vector  $\bar{\mathbf{m}}$ , and the covariance matrix  $\bar{\Sigma}$ . Based on the characteristics of the actual histogram generated from a board/image, the mean vector and covariance matrix of the ideal histogram are estimated. To compute these parameters regions in the actual histogram are selected that are believed to represent only the clearwood areas. After the parameters have been estimated using this limited data set, the ideal histogram can be extrapolated over the entire color space. Clearly, the utility of this approach is based on how well the model matches histograms of clearwood and how accurately the elements in the histogram that represent only clearwood can be found. The characteristics of the clearwood regions were first studied and verified that they indeed resembled the normal distribution and a summary of the characteristics is described in the next section. There are a number of ways to identify the clearwood regions in the histogram and these will be described in Section 5.2.

The histogram modeling technique has been implemented in 2-dimensional and 3-dimensional space. The 2-dimensional case will be described first for the sake of simplicity. Once this technique is understood it can easily be extended to three dimensions. In Chapter 6, an evaluation of the algorithms is also presented along with the results that were obtained from the many experiments that were performed. The histogram modeling technique is also compared to an existing black and white segmentation algorithm [XIAO], which is currently used in the defect detection system at the Brooks Forest Products Research Center.

## **5.1 Characteristics of Clearwood Histograms**

To develop a good model for the histogram, it is important to study the nature and shape of the histograms along with other color characteristics of wood. The regions corresponding to clearwood and defect areas have to be identified in the histogram, so that the ideal histogram for the clearwood region can be extrapolated over the entire color space. The clearwood sections of the boards have been modeled as a normal density function. To validate this model, the histograms of only the clearwood sections of several boards were studied, and one typical board is used as an example.

A large part of the wood processed in the industry is hardwood lumber that is typically represented by Oak. Other softwood categories that are often used include Yellow Poplar, Pine etc. Thus when a segmentation technique is developed it has to first work well on Oak boards. The algorithm can then be modified to be able to segment histograms of other types wood, which represent a more generalized case when compared to the Oak histograms.

Figure 5.1 shows an Oak board that is composed entirely of clearwood. It is actually a part of the clearwood section of the board shown in Figure 4.8. The histograms of this board are shown in Figure 5.2 and it can be seen that they closely resemble the normal density function. The fact that clearwood regions resemble the normal distribution and defect areas in the histogram in general do not, can be used to separate the clearwood and defect areas in the histogram, as will be explained in the next section.



Figure 5.1: Board with only clearwood regions

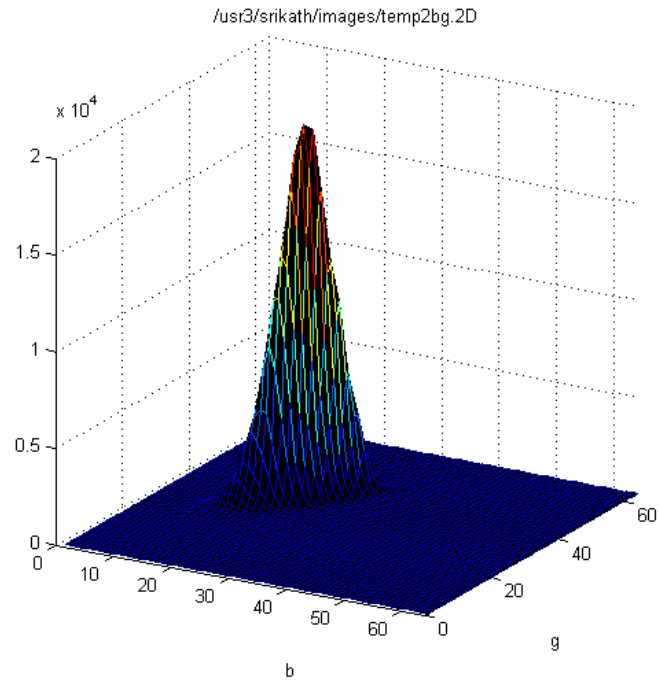


Figure 5.2 (a): b-g histogram of the board in Figure 5.1

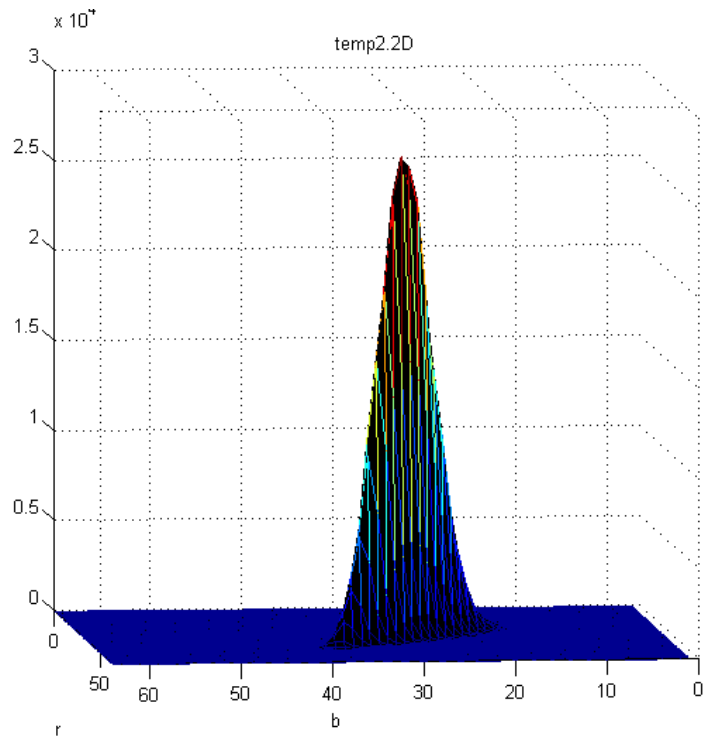


Figure 5.2(b): r-b histogram of the board in Figure 5.1

## **5.2 Developing a Model Histogram**

A large number of histograms obtained from three different species of boards (Section 3.5) were studied. For this study, a data set of boards/board images were created for each species. Each data set was created so that most every wood feature was represented on at least one board in the data set. Further, boards with varying quantities of defects on the surface were also chosen.

The general findings of the study will now be summarized with the aid of a few typical boards. Consider Figures 4.9a-b which are histograms of the board shown in Figure 4.8. It was shown in the previous section that a part of the clearwood region of the board closely resembles the normal distribution (Figure 5.2a-b). The shape of the main cluster in Figure 4.9a and Figure 4.9b also resemble a normal distribution very closely. This was the typical shape seen in histograms of boards across all species studied, when the defect area is small. However, notice how the histograms in Figures 4.9a-b deviate from the normal distribution (Figure 5.2a-b) towards the origin. This deviation is because of the dark colored knots in the board which are not present in the board in Figure 5.1. In a 2-dimensional space the darker areas are located near the origin. The knots add darker colors in the histogram that would have otherwise not been present in those quantities. The study showed that normal density function represents a good model for histograms of clearwood of a variety of species of wood.

The yellow poplar board shown in Fig. 4.4a has regions of heartwood, sap wood, and some mineral streak and other colors which are characteristic of yellow poplar. The mineral streak region appears as dark colored defect region. These regions are identified in the histogram. A schematic drawing of this histogram (Figure 4.13a) is shown in Figure 5.3, where the various regions are identified. It has to be noted that this figure is not drawn to scale, and will be used merely to explain the actual histogram. To validate the findings, the range of r,g,b values were physically examined in the heartwood and sapwood regions in the image of the board. These values were then located on the two dimensional histogram. The procedure was repeated over several boards, and one such board is used to illustrate the results. The area near the origin represents the dark colored defects. The sapwood region extends to the far end of the histogram and the heartwood region is located in the central region, which appears like a plateau. There is

however no clear visual demarcation between the two regions. In reality, there are two clusters for each region but there is extensive overlap. The sapwood region starts out like a normal distribution at the far end of the histogram and rises to a peak, but merges with the heartwood cluster after peak towards the lower end. This fact along with the presence of noise makes it impossible to find one clear peak for each cluster.

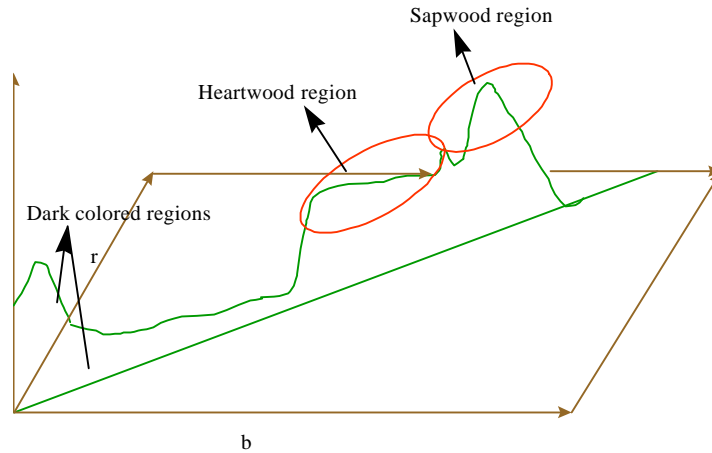


Figure 5.3: Schematic diagram of Figure 4.13a, (r-b) histogram of yellow poplar board y3a.dat

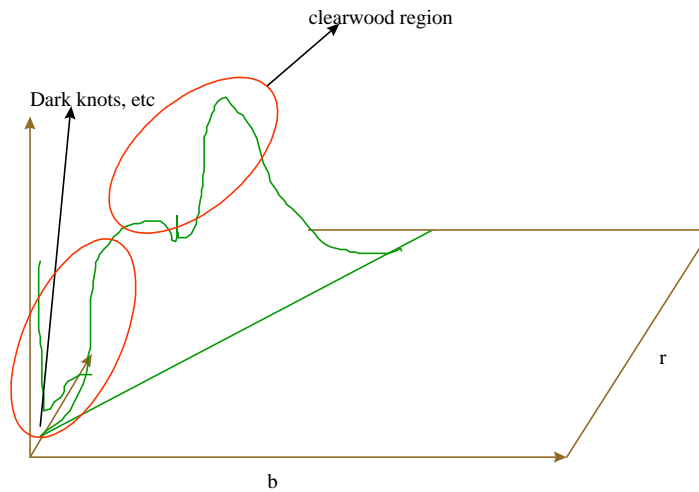


Figure 5.4: Schematic diagram of Figure 4.9a, (r-b) histogram of pine board p8.dat

A similar study of the pine boards showed the general characteristics of the clearwood regions along with the darker grain pattern. Figure 5.4 shows the schematic of Figure 4.12a, which is the histogram of the pine board shown in Figure 4.2a. The color of the grain pattern is very close to some of the brown knots and pith regions. For this reason, the multispectral clustering technique described in Chapter 4, classified some of these grain patterns as defects.

Once again, the clearwood region starts out like a normal distribution in the far end of the histogram, and is slightly distorted due to the grain pattern. It will, however, be shown that approximating the clearwood as one single distribution gives good results, provided the other parameters of the algorithm are correctly chosen. These parameters will be introduced in Section 5.3.

Once the regions corresponding to the clear-wood and defect has been broadly identified, the problem is to find suitable boundaries for these areas. Once again it is important to note that perfect separation of defects is not possible, by using only the clustering techniques. So the goal is to achieve the best possible trade off. The solution lies in developing a good model for the clear-wood regions, measuring the deviations from this model in the observed histogram and estimating a difference measurement, and finally using some form of threshold on the difference measurement to generate the map (Section 3.2) used for segmenting the image.

### **5.3 Description of the Algorithm**

The previous sections described in general the characteristics of the clearwood regions and how they resemble a normal density function. This information is used for generating a model for each histogram. The histogram of the clearwood regions is used to estimate the parameters of the gaussian distribution. These parameters are used to generate an estimate of the ideal histogram over the entire color space. A difference measurement is then used to measure the deviation of the observed histogram from the model. The difference measure must be such that the spikes in the histogram will not affect the segmentation, while maintaining the capability to detect the small and weak defect clusters. To make the algorithm robust in the presence of noise, as well as detect weak clusters, the difference measurement at a histogram element  $(i, j)$  is weighted with a number inversely proportional to the value of the model histogram at element  $(i, j)$ .

#### **5.3.1 Algorithm in 2-Dimension when Majority of the Board is Clearwood**

The algorithm is first described for the case where the board is mostly clear-wood, free from undesirable features. Here the main peak of the histogram will correspond to the clear-wood

region. This basic algorithm will be modified to handle different situations where the majority of the board need not be clearwood. The parameters of the algorithm can then be adjusted to identify a certain set of features as a defect.

The steps of the model based approach, are listed below:

1. Locate the highest peak in the histogram.
2. Take a rectangular paralleloiped region  $R$  around this point such that

$$R = p\% \Sigma H \quad \text{where,}$$

$H$  = original histogram

$p$  = Ratio of the volume of the paralleloiped to the volume of the entire histogram expressed as a percentage.

This is illustrated in Figures 5.5. The original histogram is shown in Figure 5.5a and the region  $R$ , which is 85% of the histogram (around the peak) is shown in Figure 5.5c.

3. Consider  $R$  to be a 2 dimensional normal distribution and estimate the parameters  $\bar{m}$ , and  $\bar{\Sigma}$ , where  $\bar{m}$  is a 2x1 vector of the two means  $\mu_1$  and  $\mu_2$  and  $\bar{\Sigma}$  is the 2x2 covariance matrix using the following expressions

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

$$\bar{\Sigma}_n = \frac{1}{n} \sum_{k=1}^n (\bar{X}_k - \bar{m})(\bar{X}_k - \bar{m})^t$$

$n$  = number of cells in  $R$

$$\bar{m} = \begin{bmatrix} m_1 \\ m_2 \end{bmatrix}, \quad \bar{X} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

4. Generate a normal distribution  $N$  with mean  $\bar{m}$  and covariance  $\bar{\Sigma}$  that is of the same size as  $H$

$$N(i, j) = \frac{1}{(2\pi)^p |\bar{\Sigma}|^{1/2}} \exp\left[-\frac{1}{2}(\bar{x}_{i,j} - \bar{m})^t \bar{\Sigma}_n^{-1} (\bar{x}_{i,j} - \bar{m})\right]$$

where,  $\bar{x}_{i,j}$  is a 2-dimensional column vector  $\begin{bmatrix} i \\ j \end{bmatrix}$ ,  $\bar{m}$  is a 2-dimensional column vector

and  $\bar{\Sigma}$  is the 2x2 covariance matrix. The range of  $i$  and  $j$  is from 0 - 63 since 6 bits per channel are used to generated the histograms.

5.  $N$  is a 2-dimensional normal distribution  $N(i,j)$ , where,  $i = 0 \dots 63, j = 0 \dots 63$ . Scale  $N$  to give the modeled histogram  $H_m$  such that

$$\sum_{i=0}^{63} \sum_{j=0}^{63} H_m(i, j) = \sum_{i=0}^{63} \sum_{j=0}^{63} H(i, j)$$

$$\text{where, } H_m(i, j) = N(i, j) \frac{\sum_{i=0}^{63} \sum_{j=0}^{63} H(i, j)}{\sum_{i=0}^{63} \sum_{j=0}^{63} N(i, j)}.$$

The modeled histogram obtained from region  $R$  in Figure 5.3c is shown in Figure 5.3b.

6. Find the weighted difference between  $H_m$  and  $H$  using the following expression

$$W(i, j) = \frac{|H(i, j) - H_m(i, j)|}{H_m(i, j) + \text{offset}} \quad \text{for all } i, j$$

where *offset* is a value set by the user. It is used to prevent  $W(i,j)$  from exploding when  $H_m$  is very small. The weighted difference obtained from the modeled histogram in Figure 5.5b and the original histogram in Figure 5.5a with an offset of 1, is shown in Figure 5.5d

7. Threshold  $W$  with a threshold  $Th$  to obtain  $M$  which is composed of two regions,  $C$  (clearwood), and  $D$  (defect), where region  $C$  consists of the elements of  $W$  which have a value greater than or equal to  $Th$ , and region  $D$  consists of the elements of  $N$  that have a value less than  $Th$ . The clustering in 2-dimensional color space after the weighted difference is thresholded with a  $Th$  of 60, is shown in Figure 5.5e. The entire color space is divided into two sections i.e., the defect and clearwood regions.  $M$  is the desired colormap, which can be used to map the original image into the defect and clear areas as mentioned in Section 3.2. The final segmented image is shown in Figure 5.5 (f).

The concept of a weighted difference was introduced to eliminate the effects of noise in the clear wood region. Due to the presence of noise, the observed histogram of the clear-wood section of the histogram may not be a perfect normal distribution and may have some spikes in it. If there are some irregularities in and around the peak of the histogram, their effect is greatly reduced, if the difference measurement is divided by the theoretical value (of the histogram) at

that point. Thus as we go away from the peak, the algorithm becomes more and more sensitive to any deviation from the theoretical value, which is exactly what is desired. The algorithm can ignore irregularities near the peak, but still detect the defect regions that are not very pronounced in the histogram. An offset is introduced simply to prevent the occurrence of very large numbers in the peripheral regions. One of the properties of the normal distribution is that the value never goes to zero. However, in regions away from the mean of the distribution, the values tend to be very small. This could lead to situations where  $W(i, j)$  becomes very large. This can be prevented by using a positive *offset* in the expression. The offset can either be set to 0 or some small number. The offset is also very useful in setting the sensitivity of the algorithm at the peaks relative to the periphery

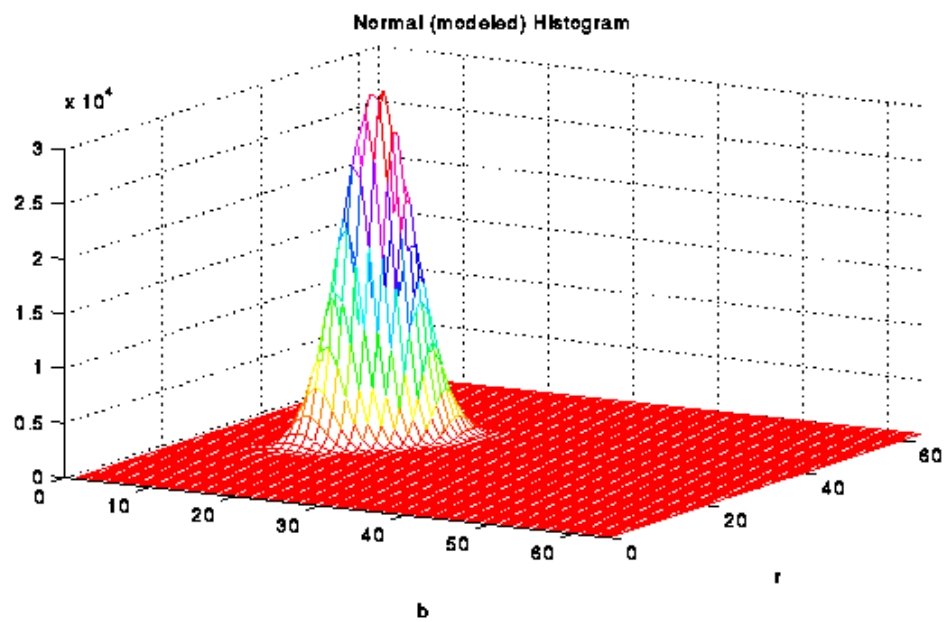
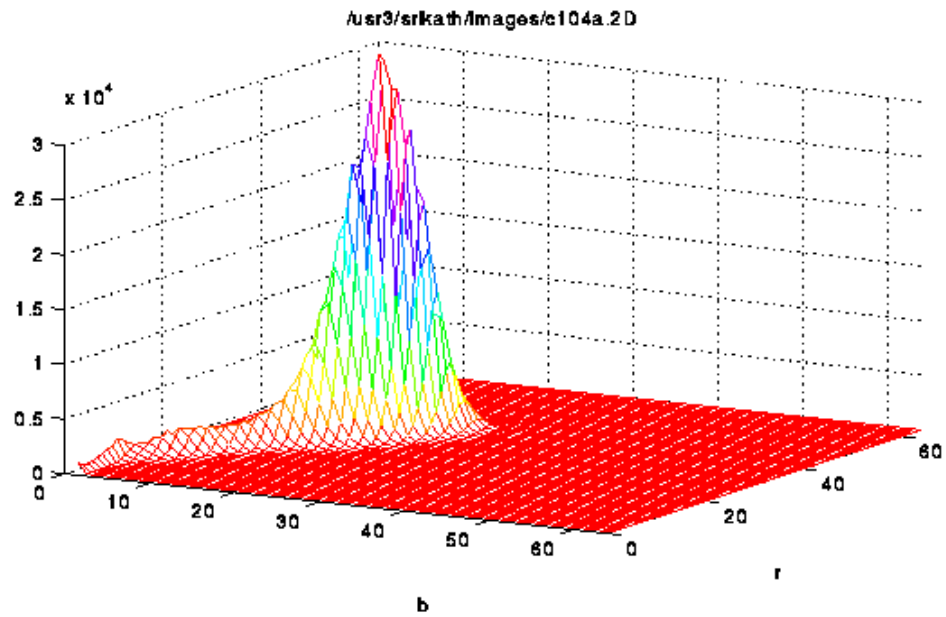


Figure 5.5: (a) Original histogram of an Oak board c104a.dat (Figure 4.5 a), (b) Modeled histogram using 85% of original histogram

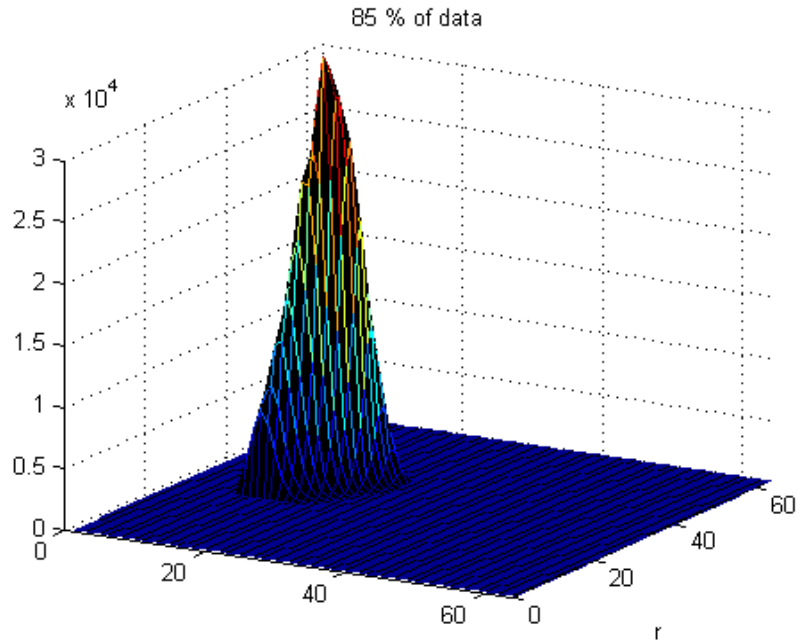


Figure 5.5: (c) Region  $R$ , 85% of original histogram

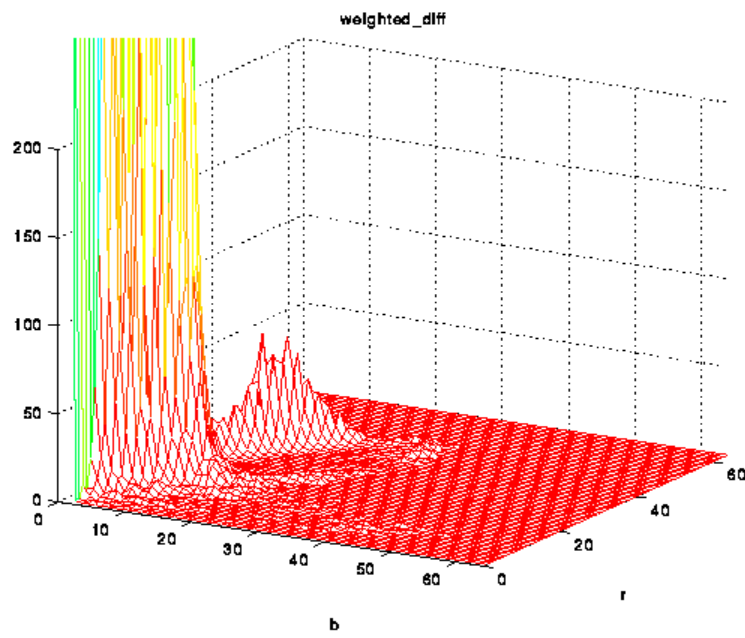


Figure 5.5: (d) Weighted difference using *offset* = 1

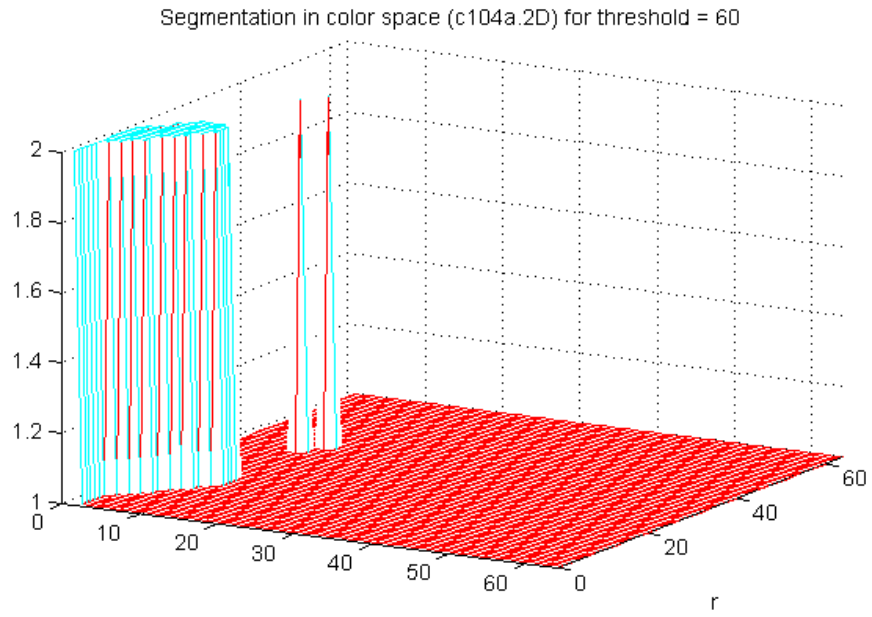


Figure 5.5: (e) Final segmentation in r-b color space using *threshold* = 60

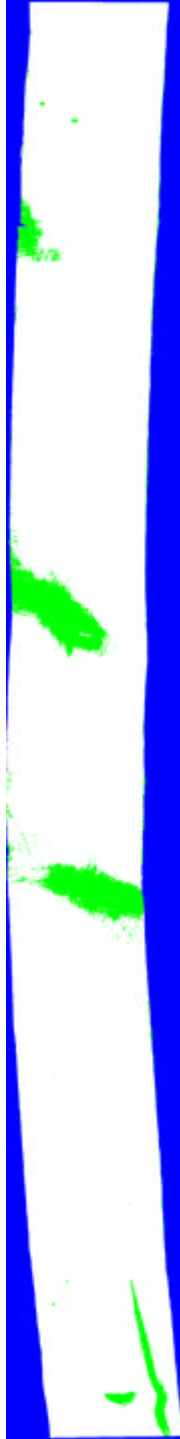


Figure 5.5: (f) Segmented image of Oak board in Figure 4.5 (a)

### **5.3.2 Extending the Algorithm**

There could be situations where the assumptions made in developing the basic algorithm in the previous section may not be entirely true. For instance, most of the boards used in commercial systems, have a majority of clearwood regions and a few defect areas. However, if blue stain is present, or there is a marked difference in the color of heartwood and sapwood regions, the majority of the board sometimes may not be clearwood. The characteristics of some of these bi-modal histograms are described in Section 5.2. Though it was possible to get good segmentation results on the boards used for the study in this Thesis, would be good to be able to apply the histogram modeling technique to 3-dimensional full color histograms for future systems. Section 5.3 explained the histogram modeling technique in 2-dimensions. In this section a more generalized approach to the algorithm is developed.

#### **5.3.2.1 Extending the Algorithm when the Majority of the Board is not Clearwood**

Now let us examine the case where the majority of the board is not free from undesirable features. One such image is shown in Figure 5.6a. The corresponding b-g, and r-g histograms are shown (Figures 5.6b, 5.6d). In this case the highest peak in the histogram is not due to clearwood, but it can be easily noted that these peaks are not in the vicinity of where clearwood generally appears. One way of locating the peak of the clearwood is by ignoring the cells in the periphery of the histogram. A peak which is located within 3-4 elements from the edge of the histogram, can be safely concluded to be from a defect cluster. Once the peak is located the rest of the process is the same as described in Section 5.2.1.

The segmented image of the pine board shown in Figure 5.6a is shown in Figure 5.6f. The defects on the board are represented by the green color. This is a most usual situation that could involve only blue stain. It can be seen (Figure 5.6(f)) that the blue stain on the original board is detected, and the clear wood regions are well separated, though the defect area is greater than the clearwood area on the board. The nature of the blue stain on the board was such that at the pixel level they did not form a continuous region on the board, though to the human eye, they seem as a large blue patch. These blue stain pixels are seen as discontinuous regions of blue stain in the

final segmented image. Other region merging techniques will have to be used to get a contiguous region in the segmented image.

### 5.3.2.2 Extending the Algorithm to 3-Dimensions

This algorithm can also be easily extended to three dimensions. The method remains the same except for the fact that all the processing is done in three dimensions. Let the original histogram of the image be  $H(i,j,k)$  where  $i, j, k$  range from  $0 \dots 63$ . Estimate the region  $R$  such that  $R = p\% H$  around the peak as explained in Section 5.3.1. Estimate the mean and covariance matrix of  $R$  as,

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

$$\bar{\Sigma} = \frac{1}{n} \sum_{k=1}^n (\bar{X}_k - \bar{m})(\bar{X}_k - \bar{m})^t$$

$n$  = number of cells in  $R$ , and

$$\bar{m} = \begin{bmatrix} m_1 \\ m_2 \\ m_3 \end{bmatrix}, \quad \text{and} \quad \bar{X} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

Generate a normal distribution  $N$  with mean  $\bar{m}$  and covariance  $\bar{\Sigma}$  which is of the same size as  $H$ .

$$N(i, j, k) = \frac{1}{(2\pi)^{3/2} |\bar{\Sigma}|^{1/2}} \exp\left[-\frac{1}{2}(\bar{x} - \bar{m})^t \bar{\Sigma}^{-1} (\bar{x} - \bar{m})\right]$$

where,  $\bar{x}$  is a 3-dimensional column vector  $\begin{bmatrix} i \\ j \\ k \end{bmatrix}$ . The range of  $i, j$  and  $k$  is from  $0 - 63$  since 6 bits

per channel are used to generate the histograms.  $N$  is a 3-dimensional normal distribution  $N(i,j,k)$ , where,  $i, j$ , and  $k$  range from  $0 \dots 63$ . Scale  $N$  to generate a modeled histogram  $H_m$  such that

$$\sum_{i=0}^{63} \sum_{j=0}^{63} \sum_{k=0}^{63} H_m(i, j, k) = \sum_{i=0}^{63} \sum_{j=0}^{63} \sum_{k=0}^{63} H(i, j, k)$$

$$\text{where, } H_m(i, j, k) = N(i, j, k) \frac{\sum_{i=0}^{63} \sum_{j=0}^{63} \sum_{k=0}^{63} H(i, j, k)}{\sum_{i=0}^{63} \sum_{j=0}^{63} \sum_{k=0}^{63} N(i, j, k)}.$$

The weighted difference  $W(i, j)$  is evaluated as

$$W(i, j, k) = \frac{|H(i, j, k) - H_m(i, j, k)|}{H_m(i, j, k) + \text{offset}}$$

The histogram of the weighted difference is then thresholded to give the final segmentation.



Figure 5.6: (a) Image of a Pine board p5.dat

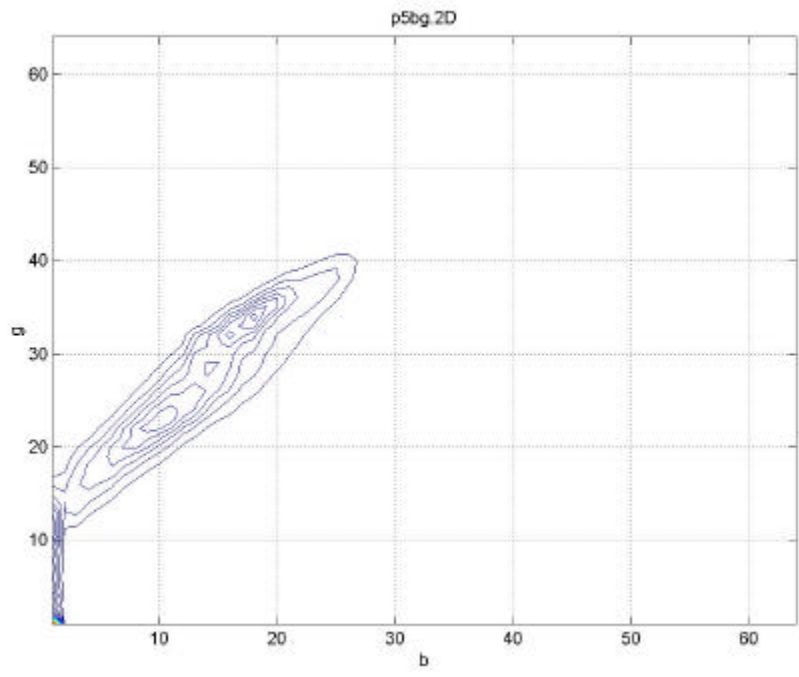
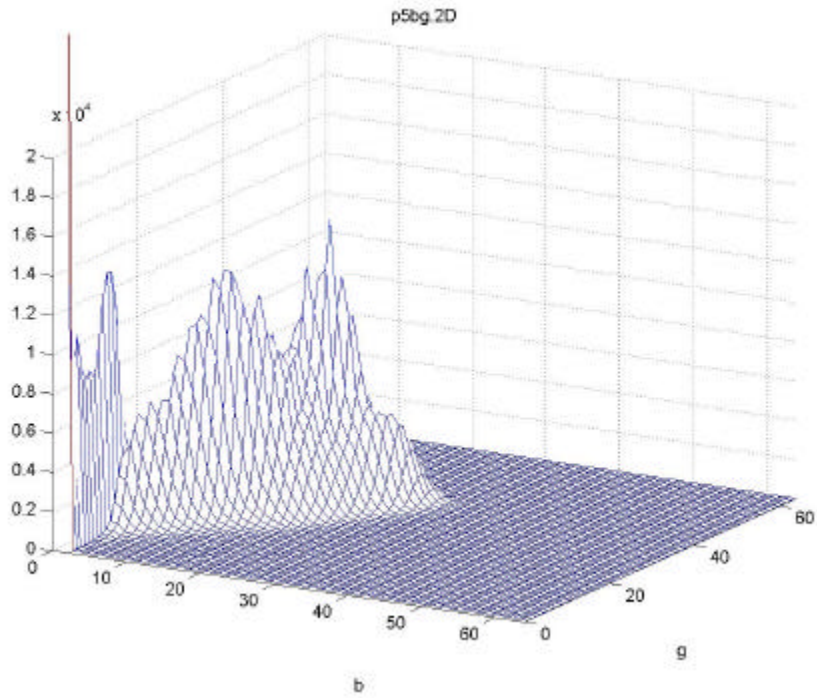


Figure 5.6: (b) Two dimensional b-g histogram of the pine board in Figure 5.6(a), (c) Contour plot of the b-g histogram

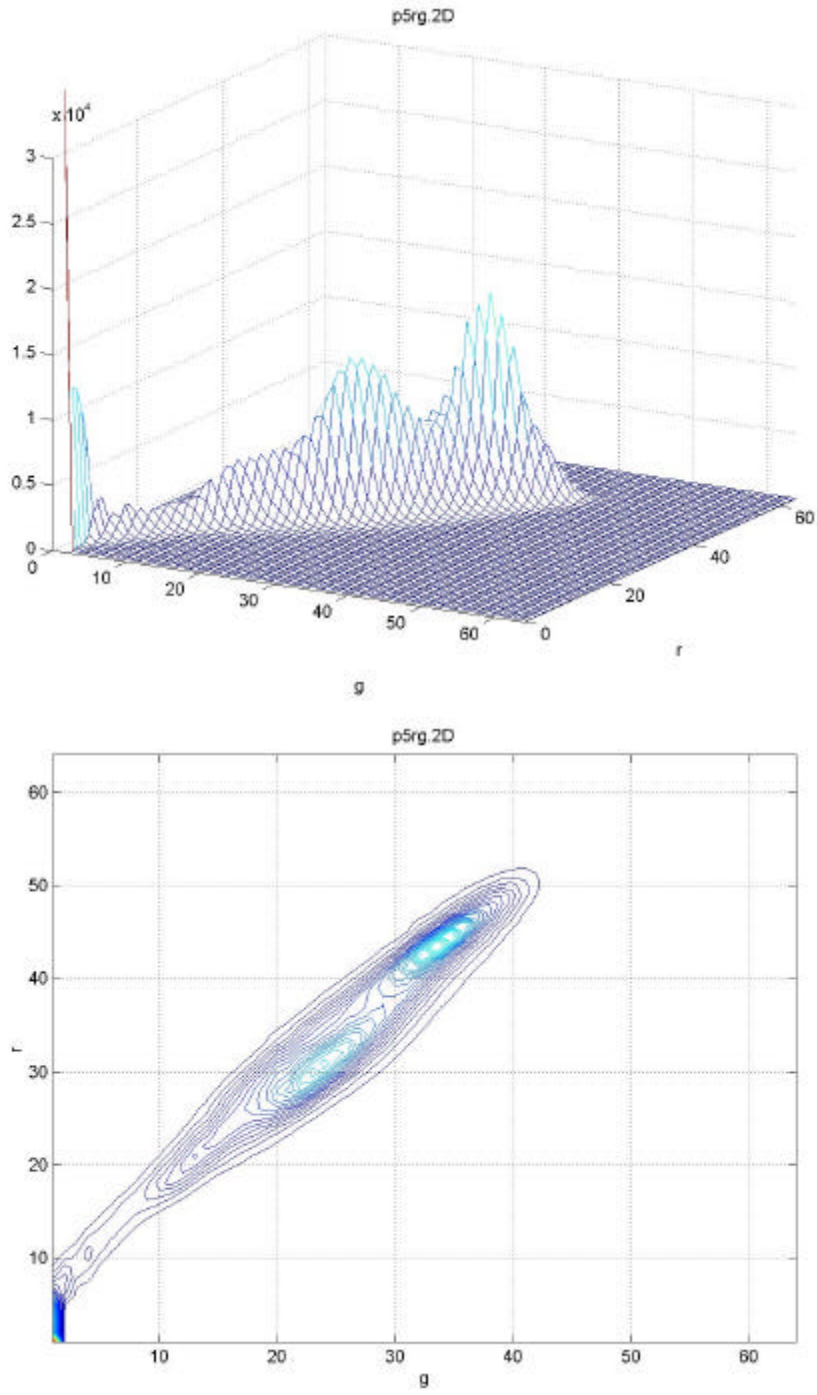


Figure 5.6: (d) Two dimensional r-g histogram of the pine board in Figure 5.6 (a), (e) Contour plot of the r-g histogram

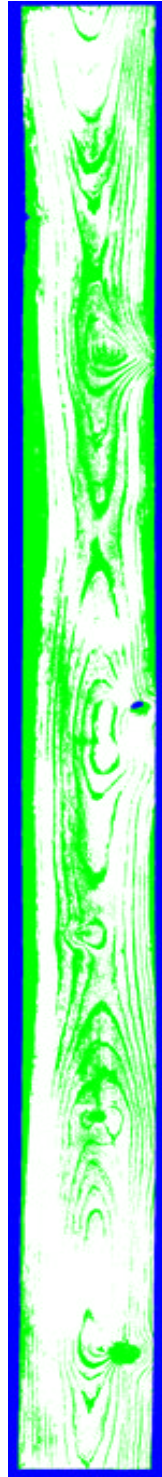


Figure 5.6: (f) Segmented image of pine board p5.dat shown in Figure 5.6a

## **5.4 Effect of the Various Parameters on the Segmentation**

The model based approach is very flexible and can be adjusted to the needs of the industry. Images of three different species of wood Oak, Pine and Yellow Poplar (Section 3.2) were scanned and processed using the histogram modeling technique. The effect of each of the parameters of the algorithm on the final segmentation are discussed below.

### **5.4.1 Number of channels used: 2-Dimensional, 3-Dimensional Histograms**

If the processing can be done using two channels to obtain satisfactory results, it is a tremendous advantage in terms of the processing time. Selecting which two channels to use is also very critical. Generally the criteria is to select those two channels which have the maximum variance for the particular species of wood. Sometimes there can be a particular wood feature that governs the selection of channels. For instance, pine boards generally have a distinct grain pattern. (Fig. 4.2a) that tend to appear as a deviation from the normal curve if the red channel is used for processing the board. All the pine boards used in the study gave good segmentation results only if the b-g histogram was used. In fact using all the three channels did not give as good results as using the b-g histogram. This will be discussed in detail in the next chapter.

### **5.4.2 Percentage of the original histogram used to create the model: $p$**

If a large region of the original histogram is used, it makes the algorithm less sensitive to features on the board. A lower value of  $p$  needs to be used when tighter constraints are to be placed on the clearwood region. This can be illustrated with an example. Consider the yellow poplar board shown in Fig. 4.4a, described in Section 4.2. The histogram of the red and blue channels have three main regions corresponding to the three areas namely heartwood, sapwood, dark colored mineral streak (Figure 5.3). Say, for an application only the sapwood areas are acceptable, and every other feature in the board including heartwood is to be treated as unsuitable. Taking a small percentage of the original histogram, would be advisable. This would give the modeled histogram a better fit for the sapwood area and the difference measurement can be used to separate the other areas. On the other hand if there is no distinction

to be made between heartwood and sapwood, a larger area has to be taken into consideration while creating the modeled histogram.

Experimentation with all the yellow poplar boards, showed that values of around 40% was generally useful for separating heartwood from sapwood, while values greater than 60% did not cause any distinction between them. These values apply to the typical boards which are seen in yellow poplar (and were used in the study), and also boards where the majority of the board consists of sapwood. If the majority of the board consists of heartwood, and very small quantities of sapwood is present,  $p$  might have to be less than 40%. However, if the peak is located in the correct location (sapwood region), the error will be greatly minimized, since the mean of the histogram is still estimated correctly. Locating the appropriate region from where the peak is obtained is discussed in the next Section. Thus the main peak (mean) in the modeled histogram will correspond to the sapwood region in the histogram.

If the sapwood area is much larger than the heartwood area, and  $p$  is set to 40%, it is possible that not all the sapwood region on the board is being used to model the histogram. However, using a part of the sapwood area should still give correct results, since the sapwood area of the original histogram that was not used to develop the model, has the same shape as that of the model.

The effect of a particular value of  $p$  is not entirely independent of the other parameters. For instance a lower threshold can sometimes compensate for taking a large region in the original histogram, but if the histogram becomes larger than a certain size, reducing the threshold will not cause the desired separation.

### **5.4.3 Selecting the region from where the peak of the histogram is obtained**

The region (in the histogram) where the peak is located becomes relevant in two situations. First, if a large portion of the board is defective and the clearwood region that has to be identified consists of only a small portion of the board surface. Second, if only certain specific feature/features on the board have to be detected, features that may not be the clearwood area,

then the required region alone can be modeled and everything else present on the board can be ignored.

The most common situation that was encountered in the boards used for testing was the presence of a large area of mineral streak, decay, wane or blue stain. In the case of mineral streak, decay and wane, the colors corresponding to these regions are present near the origin. The regions in the histogram corresponding to blue stains are not present near the clearwood cluster. In both these situations (boards with dark colored defects like mineral streak, wane etc., and blue stain) ignoring the histogram at all four edges of the color space ( $i, j = 0-2$  or  $60-63$ ), while locating the peak, is sufficient to give good results.

#### **5.4.4 Value of the threshold $Th$ used on the Weighted histogram $W$**

The value of  $Th$  is directly related to how sensitive the algorithm will be in detecting defects. A low value of the threshold makes it more sensitive to small deviations from the modeled histogram. Experimentation showed that the final segmentation is unaffected by a 5% change in the value. The value of threshold is also closely related to the value of  $p$  used in the clustering process. Using a small value of  $Th$  can reduce the effect of using a large value of  $p$ , and vice-versa.

#### **5.4.5 Value of Offset**

The purpose of *offset* is to prevent the weighted difference from getting very large. There is a possibility of this happening in regions away from the mean of the modeled histogram ( $H_m$ ), where the values of  $H_m$  can be very low. Using a larger value of offset has the effect of increasing the threshold. However more extensive testing on a large number of samples needs to be performed, to examine if this has any other consequence on the segmentation. This is beyond the scope of this work.

## **5.5 Conclusions**

A study of the applicability of conventional clustering techniques on wood histograms has not been performed previously. Based on such a study performed in Chapter 4, the concept of a model based approach to clustering was developed. Experiments showed that the clearwood regions of wood can be modeled as a gaussian distribution. A gaussian model of the histogram is developed based on part of the histogram around the clearwood peak. Deviations from this model are used to determine the defect areas. The concept of a weighted difference was introduced. The weighted difference makes the difference measurement immune to noise around the peaks of the histogram, but sensitive to even small deviations from the model in regions far away from the mean. Thus the weak defect clusters are identified even though the algorithm is robust to noise. A qualitative analysis of the various parameters of the algorithm has also been discussed. The results of applying the model based approach to actual histograms and a comparison of the technique to a b/w segmentation algorithm in current use will be discussed in the next chapter.