Chapter 5

Feature Classification Using Color, Shape, and Density Parameters

5.1 Introduction

Currently, wood feature classification methods use the parameters that are provided by a particular sensor type as classification variables. No knowledge is gained before developing a classifier about how feature types are represented by the sensing technology. It is hypothesized that by understanding how wood features are represented by the parameters measured, and understanding the relationship of these parameters, that the best possible classification method may be developed. In this Chapter the relationships of parameters determined in Chapter 4 were used to develop discriminant classifiers, and to demonstrate that with a knowledge of how features are represented by color, shape, and density parameters the best possible classification accuracy can be achieved for wood features. By comparing the accuracy of classifiers, the effectiveness of certain parameters as classification variables was determined. The classification parameters were selected based on the knowledge of parameter relationships developed in Chapter 4 and by using forward, backward, and stepwise variable selection procedures. Discriminant classifiers were developed to differentiate features within species, between species, and compare the abilities of RGB and HSI color parameters to classify feature types.

By gaining knowledge of how wood features are represented by color, shape, and density parameters, it is hypothesized that better classification methods can be developed. The improvement in classification is based on selecting the parameters that provide good separation variables in the classifier, while leaving out poor ones. In Chapter 4 the relationship of parameters between feature types was determined using the significant differences of parameters between features. The parameters that were compared and their notations are listed in Table 4.1. Parameters that showed significant differences between features were considered good for feature differentiation methods, and those that had few or no differences were considered poor. When multiple parameters give the same differences between feature types often one can be left out of the classifier. Using both variables will not add to the accuracy of the classifier, and may actually reduce its performance (Everitt and Dunn, 1992).

5.2 Discriminant Analysis Defined

Discriminant analysis is one method available to develop functions that can classify unknown groups based on known groups. Discriminant analysis is a statistical technique that is used to classify objects into groups based on *a priori* information. Discriminant functions are often used in pattern recognition and classification for the separation of data into groups (Duda and Hart, 1973).

The most common methods of discriminant analysis include Fisher's Discriminant function and linear logistic discrimination (Everitt and Dunn, 1992). Other methods are available and are reviewed in Hand (1981). The many different methods are a result of the variety of distributional assumptions made about the variables describing the feature to be classified, however, Fisher's linear discriminant function has been shown to be relatively robust for those situations where there is a departure from normality (Everit and Dunn, 1992). Fisher's approach to constructing a classification rule is based on specifying a theoretical probability distribution model (normal) and assuming that the data fits the model, then estimating the parameters using the data, and finally constructing a rule using these estimates (Everit and Dunn, 1992).

Discriminant classifiers can be broken into two classes, those where classification criterion is based on the individual within-group covariance matrices, and those based on the pooled covariance matrix required by the homogeneity of the within-group covariance matrixes. The first group of classifiers is known as quadratic classification functions. In general, each observation is placed in the class from which it has the smallest generalized squared distance. For example, a feature f is assigned to feature class (class) c if

$$Q_{fc} > Q_{fc'} \qquad \textbf{Equation 5.1}$$

where Q_{fc} is the quadratic classification function score for feature f in class c, and $Q_{fc'}$ is is the quadratic classification function score for feature f in any other class c. The quadratic classification function score is calculated as

$$Q_{fc} = \ln q_c - \tfrac{1}{2}\ln\left|S_c\right| - \tfrac{1}{2} D^2_{fc} \qquad \textbf{Equation 5. 2}$$

where q_c is the estimated prior probability for class c, S_c is the covariance matrix for feature class c, and D^2_{fc} is defined as sample distance between feature f and class c.

When the within-group covariance is used, the distance from feature f to feature class c is

$$D^2_{fc} = \left(X_f - \overline{X}_c\right)' S_c^{-1}\left(X_f - \overline{X}_c\right) \qquad \textbf{Equation 5. 3}$$

where, D^2_{fc} is the sample distance between feature f and class c, S_c is the covariance matrix with feature class c, X_f is a p-dimensional vector containing the quantitative variables of an feature, X_c is the p-dimensional vector containing the variable means in class c.

The misclassification rate of the discriminant function can be determined several ways as described in Hand (1986). The most common method is the "leaving one out method" where the discriminant function is derived on the basis of (n-1) of the subjects and is used to classify the individual not included (Everitt and Dunn, 1992). The whole process is then repeated for each individual.

Discriminant analysis also allows a reduction in variables to be achieved without effecting the classification error rate. Several different procedures are available to determine if all the variables are required to provide a correct classification. The general principle with all these procedures is to choose a measure of separability between the groups and then sequentially eliminate those groups whose removal leads to the least reduction in separation (Hand 1981). Two common procedures include the stepwise and backward variable selection methods. The stepwise analysis begins with no variables in the model. At each step the model is examined and if a variable entered fails to meet the Wilk's lambda criterion (Huberty, 1994) then it is removed. When all variables entered in the model meet the criterion, then the stepwise method stops. The backward selection method begins with all the variables in the model and at each step the variable that contributes the least to the discriminatory power of the model, as measured by Wilk's lambda, is removed. The process is stopped when all the remaining variables meet the criterion to stay in the model.

The number of samples required for each feature class to remain within 0.05 of the optimum error rate is approximately 2k when δ is large and 3.5k when δ is small, where k is the number of classes and δ is the calculated distance of known parameters using equation 5.3. (Lachenbruch, 1974). The number of samples collected for each feature type is listed in Table 3.4. The largest number of classes in the selected parameter discriminant classifiers is seven. This number falls with the acceptable level of samples required. However, some of the discriminant classifiers compared against contained up to 12 variables. Usually the larger number of variables creates overfitting of the data and higher classification accuracies would result. Since none of the tested classifiers were shown to have higher classification accuracies than those with selected parameters, the methods used are presumed sound. Also, it must be recognized that in determining the classification accuracy of the classifiers the leave one out method was used where one feature type is compared against the total sample size.

5.3 Variable Selection for Discriminant Classifiers

Discriminant functions, or classifiers, were developed using the measured parameters to classify each feature type and to demonstrate the marginal importance of each parameter in classification. The selection of parameters (variables) to be included in a discriminant classifier can have great impact on the classification results. The reduction of the total number of parameters in the discriminant classifier is important because the fewer the predictors relative to the number of samples, the more accurate and more precise the estimators (Huberty, 1994). Variables that do not contribute to the classification of features can also be removed. Variable selection methods based on group separation can be used to reduce the total number of parameters in the discriminant classifier, and as a method to demonstrate the importance of certain parameters in differentiating between features. The criterion for adding or deleting variables is based on group separation (Huberty, 1994). As mentioned earlier, common methods of variable selection in discriminant analysis include forward, backward, and stepwise (Huberty, 1994 and Everitt and Dunn, 1992). While variable selection methods provide a way to reduce variables when little is known about the data, the selection of parameters is better when based on previous research and knowledge of what parameters would provide good classifiers. Statistical screening using multiple univariate analysis (ANOVA) can often provide this knowledge (Huberty, 1994).

As discussed above, the parameter relationships for wood features discovered in Chapter 4 were used in selecting the appropriate parameters for classification and were compared with those selected using forward, backward, and stepwise procedures. Comparisons between sets of selected parameters was based on the accuracy of the discriminant classifiers, where accuracy is defined as having a low number of misclassifications in each feature group. The criterion for feature classification was based on the within-group covariance matrices, S_c. Each feature to be classified was placed in the feature class according to equations 5.1-5.3.

All developed classifiers were tested using the leave-one-out method discussed in section 5.2 rather than testing new data sets. This is an accepted method for validating the ability of discriminant classifiers (Huberty, 1994 and Everitt and Dunn, 1992).

5.3.1 Single Variable Discriminant Classifiers

Discriminant classifiers were first developed and tested for each parameter separately to demonstrate the effectiveness of these parameters as classification variables. It was predicted in Chapter 4 that variables, which have significant differences between many feature types, will provide good classification variables in discriminant classification as suggested by Huberty (1994). The classification results are presented for each species in Table 5.1-Table 5.3.

For red oak it was predicted in Chapter 4 section 4.2.1 that R_m, G_m, B_m, H_m, and I_m, would make good classification variables and that X_m, ASP, I_s, and X_s would make good classification variables when used in combination with other parameters. The overall classification accuracies seen in Table 5.1 for the R_m, G_m, B_m, H_m, and I_m parameters are the highest overall, suggesting that the methods used to select classification variables in Chapter 4 are appropriate. The classification accuracies of X_m, ASP, I_s, and X_s were not high as predicted. However, this does not mean these variables are not useful in a multi-parameter classifier. Testing the ability of these parameters in a multi-variable classifier was conducted in the next section.

The parameters selected as good classification variables for hard maple in section 4.2.2 were R_m, G_m, B_m, I_m, X_m, ASP, X_s, and S_s. The H_m, ASP, X_m, and X_s were predicted to be good classification variables in a multi-parameter classifier. In Table 5.2 it can be seen that the overall classification accuracy of R_m, G_m, B_m, I_m, and S_s is very high as predicted and that the accuracies for H_m, ASP, X_m, and X_s are moderate to low. For white pine R_m, G_m, B_m, and I_m, were recommended as good classification variables and R_s, S_m, ASP, RND, X_m, and X_s were recommended as good classification variables for multi-variable classifiers. The classification results seen in Table 5.3 for the R_m, G_m, B_m, and I_m, in white pine were high and R_s, S_m, ASP, RND, X_m, and X_s were moderate to low. The results of classification accuracies in the single parameter discriminant classifiers helps to demonstrate the ability of the variable selection

Table 5.1. Classification results for each parameter for red oak features.

	Classification Accuracy (% classified correct)				
Parameter	**Bark**	**Clearwood**	**Knots**	**Stain**	**Overall**
R_m	95	100	65	21.05	70.26
G_m	95	100	60	21.05	69.01
B_m	95	100	60	21.05	69.01
X_m	20	65	45	0	32.50
ASP	45	10	90	47.37	48.09
RND	65	0	95	52.63	53.16
R_s	40	50	60	0	37.50
G_s	0	75	85	15.79	43.95
B_s	20	85	80	5.26	47.57
X_s	20	80	50	31.58	45.40
H_m	40	85	10	42.11	49.28
S_m	45	95	10	52.63	50.66
I_m	95	100	55	21.05	67.76
H_s	65	95	55	0	53.75
S_s	90	55	30	25	50.00
I_s	0	75	75	15.75	41.44

Table 5.2. Classification results for each parameter for hard maple features.

	Classification Accuracy (% classified correct)				
Parameter	**Bark**	**Clearwood**	**Knots**	**Stain**	**Overall**
R_m	100	95	80	68.75	85.94
G_m	100	95	80	56.25	82.81
B_m	92.86	95	80	18.75	71.65
X_m	14.29	70	55	62.5	50.45
ASP	28.57	0	90	31.25	37.46
RND	21.43	40	90	18.75	42.55
R_s	57.14	90	25	43.75	53.97
G_s	42.86	80	25	37.5	46.34
B_s	14.29	80	25	31.25	37.64
X_s	50	85	35	0	42.50
H_m	71.43	55	60	56.25	60.67
S_m	71.43	95	25	56.25	61.92
I_m	100	95	80	56.25	82.81
H_s	64.29	70	70	62.5	66.69
S_s	78.57	90	60	56.25	71.21
I_s	42.86	90	25	37.5	48.84

Table 5.3. Classification results for each parameter for white pine features.

	Classification Accuracy (% classified correct)				
Parameter	**Bark**	**Clearwood**	**Knots**	**Stain**	**Overall**
R_m	100	100	65	76.92	85.48
G_m	100	100	65	84.62	87.41
B_m	100	100	60	92.31	88.08
X_m	44.4	0	85	15	36.10
ASP	44.4	0	80	7.69	33.02
RND	33	0	95	30.77	39.69
R_s	0	90	70	7.69	41.92
G_s	0	75	65	0	35.00
B_s	44.4	5	15	0	16.10
X_s	55.6	15	60	76.92	51.88
H_m	33.3	85	85	15.38	54.67
S_m	44.44	85	70	0	49.86
I_m	100	100	65	84.62	87.41
H_s	44.44	100	85	38.46	66.98
S_s	66.67	90	65	0	55.42
I_s	0	80	65	0	36.25

technique suggested by Huberty (1994) and used in Chapter 4. Further verification of these parameters for classification will be discussed for multi-parameter classifiers in the next section. It is interesting to note that stain/mineral streak had the lowest overall classification accuracies for the feature types. This can be explained by the lack of parameters that were significantly different for stain/mineral streak and all other feature types as shown in Chapter 4.

5.3.2 Multi-variable Discriminant Classifiers

Discriminant classifiers were developed for each color space and species using multiple parameters. The discriminant classifiers, for all species and color spaces, developed using parameters selected by the forward, backward, and stepwise methods all proved to have lower classification accuracies than those developed based on parameter relationships determined in Chapter 4. Also, in all cases the forward and stepwise methods produced the same results; therefore, only the forward and backward selection methods are presented and discussed. The parameters selected by each method are presented in Table 5.4 and Table 5.5. The development of discriminant classifiers using parameters selected based on Chapter 4 results is continued in the following sections based on the color space used.

5.3.2.1 RGB Color Space Parameters

The parameters selected by each method are presented for each species in Table 5.4. The selection of parameters for the RGB based classifier was determined from statistical relationships discussed in sections 4.2. R_m, G_m, and B_m were indicated in section 4.2 to have the same significant differences between feature types. Further evidence for supporting these differences was provided by the similar classification results for all features except stain as shown in Tables 5.1-5.3. Based on ANOVA results the R_m, G_m, and B_m parameters appear to give equal classification results. However, there are other relationships that can be used to determine the appropriate parameters for classification. R_m was chosen as a classification variable because it has a larger magnitude in the differences between features. B_m was chosen because it had one more significant difference between feature types for all species than G_m.

Table 5.4. The RGB color based parameters selected for each species by the forward method, (F), backward method, (B), and from statistical relationships determined in Chapter 4, (S).

Species	Selection Method	Selected Parameters									
Red Oak	F & B	R_m	G_m	X_m	ASP	RND	R_S	G_S	B_S	X_S	
	S	R_m	B_m	X_m	ASP	R_S	B_S	X_S			
Hard Maple	F	R_m	G_m	B_m	X_m	RND	ASP	R_S	X_S		
	B	R_m	G_m	B_m	X_m	RND	ASP	R_S	G_S	B_S	X_S
	S	R_m	B_m	X_m	ASP	R_S	X_S				
White Pine	F	R_m	G_m	X_m	ASP	R_S	B_s	G_S	X_S		
	B	R_m	G_m	X_m	ASP	B_S	G_S	X_S			
	S	R_m	B_m	X_m	ASP	R_S	X_S				

Table 5.5. HSI color based parameters selected for each species by the forward method, (F), backward method, (B), and from statistical relationships determined in Chapter 4, (S).

Species	Selection Method	Selected Parameters									
Red Oak	F &B	I_m	X_m	ASP	RND	H_S	S_S	I_S	X_S		
	S	H_m	I_m	X_m	ASP	I_S	X_S				
Hard Maple	F & B	H_m	S_m	I_m	X_m	ASP	RND	H_S	S_S	I_S	X_S
	S	H_m	S_m	I_m	X_m	ASP	S_S				
White Pine	F	I_m	X_m	RND	H_S	S_S	I_S	X_s			
	B	I_m	X_m	ASP	H_S	S_S	I_S	X_S			
	S	S_m	I_m	X_m	ASP	S_S	X_S				

This combination of color mean parameters was tested against all possible combinations and it was discovered that the G_m and B_m parameters could be interchanged without affecting the classification accuracy of features in red oak and hard maple. G_m, however, provided higher classification accuracy than B_m in white pine. The ability to interchange G_m and B_m in the discriminant function should have been expected, as the p-values for these parameters presented in Tables 4.3 and 4.4 were equal between the various feature comparisons. It was also found that the R_m consistently provided higher classification accuracies than G_m or B_m for all species. B_s was added to red oak because of the lack of other parameters which were significantly different between knots and stain/mineral streak.

When all discriminant classifiers were compared, those with two color mean parameters were found to out perform those with only one or three, indicating that two color mean parameters are required for the highest classification accuracy. These findings are similar to those found in Conners et al. (1985), Maristany et al. (1991), and Brunner et al. (1992). Both shape and density parameters were found to increase performance of the discriminant classifiers. By comparing the classification accuracies of numerous discriminant functions without shape measures to those with shape measures, it was discovered that including a shape measure increased the classification accuracy for all feature types, except clearwood, by 5-10%. The clearwood classification accuracy was not improved because the clearwood shape parameters were an average of all the feature type shape parameters and were included only so the discriminant classifier would not be biased to the clearwood shape measures. The greatest increase in classification results for shape measures was for bark pockets. By comparing classifiers with both ASP and RND it was determined that ASP was a better classification parameter than RND in red oak and hard maple. Using ASP instead of RND attained a 5% increase in the classification of bark pockets. Both shape measures were found to be equal for white pine feature classification. The contribution of density to classification varied for each species, however, density improved the classification of stain by 5-10% for all species. While density alone does not provide high classification accuracies for feature types as seen in Tables 5.1-5.3, when included in a multi-parameter model it was found to increase the detection of knots in hard maple and clearwood in red oak by 5%.

By comparing the classification accuracies for each species specific classifier, as seen in Table 5.6, it can be concluded that by using knowledge about how features are represented by color, shape, and density parameters and knowing how those parameters are related, a better overall classification function can be created.

While the overall classification accuracy for the selected variable discriminant classifier provided accuracies of 97.3 % or better, some feature types proved more difficult to classify than others. The feature type most difficult to classify varied for each species. However, the most confusion involved stain/mineral streak. Red oak had difficulty with bark pockets and knots, both of which were missclassified as stain/mineral streak. This difficulty can be explained by the three classification parameters (R_m, B_m, and R_s) that were not significantly different between these feature types as seen in Table 4.3. The feature type most often misclassified in hard maple was clearwood that was misclassified as stain. In white pine, stain was misclassified as clearwood. Stain has also proved a difficult feature to classify by other investigators, due to the large variability in the feature and color similarity with other features (Adel et al., 1993).

The discriminant classifiers that provided the highest classification accuracies used less variables than those selected using forward and backward variable selection methods. The classifiers that used all parameters in the discriminant classifier had the lowest accuracies. The reduction of the total number of parameters in the discriminant classifier is important because the fewer the predictors relative to the number of samples, the more accurate and more precise the estimators (Huberty, 1994).

Table 5.6. RGB color based classification accuracies for species specific classifiers where F is the forward method, B is the backward method, A is all parameters, and S for parameters selected from parameter relationships determined in Chapter 4.

Species	**Variables Selected**	**Feature Classification (% classified correct)**				
		Barkpocket	**Clearwood**	**Knot**	**Stain/Mineral**	**Overall**
Red Oak	F	100	95	85	100	95.0
	B	90	95	90	89.5	91.1
	A	95	95	90	94.75	93.7
	S	95	100	95	100	97.5
Hard Maple	F	100	95	95	88	94.5
	B	100	95	95	93	95.8
	A	100	95	95	81.25	92.8
	S	100	95	100	94	97.3
White Pine	F	0	95	100	84.62	69.9
	B	78	95	100	84.62	89.4
	A	0	95	100	84.62	69.9
	S	100	100	100	92	98.0

5.3.1.2 HSI Color Parameters

The same method for selecting model variables for the RGB classifiers was applied to parameters based on the HSI color space (see Table 5.5) . The classification accuracies of each developed species-specific discriminant function are listed by each feature type and for the overall classification accuracies in Table 5.7.

The classifier using variables selected from statistical analysis of parameters outperformed all other classifiers as noted in Table 5.7. The variables were selected using the following reasoning. I_m was included in all the classifiers because it was found to have the largest number of significant differences between feature types for all species. I_m had more differences between feature types than H_m or S_m for all species. Tables 5.1-5.3 also show that the I_m provides good classification accuracy for all feature types except stain/mineral streak when used as the only parameter in a discriminant classifier. The significant differences for the H_m and S_m parameters for features varied greatly between species and the effectiveness of these parameters in the discriminant classifier also varied with each species. H_m was used in red oak because there were more significant differences between feature types than for S_m. S_m was chosen for hard maple because larger difference in magnitude between feature mean values. Results of the classifier using these selected parameters were higher than all others, as seen in the overall classification accuracies listed in Table 5.7, confirming the ability of the ANOVA method of variable selection.

It was noticed when testing the various combinations of classification parameters for white pine that the I_m could be used as the only color parameter and perform equal to the model that included the S_m. When the I_m is used as the only color parameter in red oak and hard maple, the classification accuracies were reduced by 15-20% for each feature. The I_m provides the same grayscale information that a black and white camera would provide. The importance of color in the detection of hardwood features has long been suggested. These results verify that color information is indeed more important in the classification of hardwood

Table 5.7. HSI color based classification accuracies for each species specific classifier where F is the forward method, B is the backward method, A is all parameters, and S for parameters selected from parameter relationships determined in Chapter 4.

Species	**Variables Selected**	**Feature Classification (% classified correct)**				
		Barkpocket	**Clearwood**	**Knot**	**Stain/Mineral**	**Overall**
Red Oak	F & B	90	95	75	94	88.5
	A	85	95	75	89.5	86.1
	S	90	100	90	100	95.0
Hard Maple	F & B	100	95	95	93.75	95.9
	A	100	100	95	87.5	95.6
	S	100	95	95	100	97.5
White Pine	F	77.8	100	95	100	93.2
	B	77.8	100	95	100	93.2
	A	0	95	90	76.92	65.5
	S	100	100	100	100	100.0

features than those features in pine species. This result can be attributed to the magnitude in the differences between features for these species as noted in section 4.2. While red oak features have less variability within, the variability between features is greater than variability in white pine features as discussed in section 4.3.1.3. Hence, more parameters were required for classification.

When classifiers with shape and density parameters were compared to classifiers without shape and density parameters it was determined that similar to the RGB model, shape and density parameters both proved to increase the accuracy of the classifier for all species. The hard maple classifier benefited more from the shape parameter than did red oak or white pine. The contribution of density to classification accuracy varied between species and features, in oak there 5% increase in bark pocket and stain/mineral streak detection, maple had an 8% increase in stain/mineral streak detection, and in white pine classification of knots increased 10%.

As with RGB color based models, most misclassifications occurred with stain/mineral streak. The lack of classification accuracy for stain/mineral streak can be accounted for by the lack of parameters that provide good classification results for this feature type in hardwoods as noted by the low classification accuracies of the single parameters as noted in Tables 5.1-5.3. A combination of parameters is required to accurately differentiate stain from other features as can be noted by comparing the classification accuracies of stain/mineral streak for single parameter classifiers in Tables 5.3-5.5 to the multiple parameter classifiers in Tables 5.6 and 5.7.

5.3.4 Summary of Classification Variables

The increase in classification accuracy due to the selection of proper variables demonstrates that having an in-depth knowledge of how parameters measured from wood features relate to one another within species, allows the best possible classification functions to be developed. Wood feature classifiers were developed based on the knowledge of parameter relationships. These classifiers were shown to outperform those where all parameters were used and those where stepwise variable selection procedures were used. Parameters that provided the best separation between feature types were included in the classifier and those that were found not to contribute were removed. The knowledge gained about parameter relationships was also used to explain classification errors. It can be concluded that for hardwoods, regardless of the colorspace, two color mean parameters are required for classification. Also, shape and density parameters were demonstrated to improve classification accuracies. It must be noted that the classification accuracy obtained in these results is fairly high. This is in part due to the classification being based on manually selected feature regions. If regions were segmented based on other methods, it is possible that more variability would exist and lower classification accuracies occur. Also, the feature classes which had their regional parameters measured were grouped based on anatomical and parameter similarities, thus reducing variability within classes. The intent of this study was to remove the variability associated with segmentation, so that the interaction of scanning parameters and classification rates could be analyzed for various species and feature types. In the industrial environment feature classes will include more variation.

To demonstrate that the effect of having greater variation in the feature classes, the subclasses of features created in section 3.6.3 for red oak were combined to form new feature classes as defined in Table 5.8. These new feature classes were classified using the parameters listed Tables 5.4 and 5.5. The classification accuracies for those feature classes are listed in Table 5.9. The classification accuracy for the combined feature classes is significantly lower for all features except clearwood for both colorspaces. It can be concluded that the increased variation resulting from mixing feature subclasses does reduce the classification accuracies of the features. The greatest number of misclassifications occurred when features were classified as stain/mineral streak when they were not. It was noted in section 5.3.2.1 and 5.3.2.2 that most misclassifications were based on stain/mineral streak. A possible limitation to this comparison is that the classification model used the selected

Table 5.8. Feature classes formed from subclasses for red oak.

Feature Class	Previously Defined subclass	Description of Feature	Differences between features
Knots	Knot_A	Intergrown knot with some checking. No decay is permitted. The branch growth is intergrown on both faces	Uniform density and color properties
	Knot_B	Intergrown knot on the scanned face; however, there is no presence of branch growth on the opposite face.	Density should be lower due to less higher density knot material.
	Knot_C	Intergrown knot with no decay on scanned face. The opposite face contains decay.	Density should be lower due to less dense material within the knot. Color properties should appear darker due to stain.
Bark Pocket	True Bark Pocket	An area which contains bark material that is the result of damage of some type where the cambium has overgrown and included bark in the wood material. Is not an area associated with an old encased knot or loose knot.	Higher density, more uniform color properties
	Bark Pocket / Encased Knot	Scanned face contains very dark stained and decayed material. There is 75% material on face. No holes. May or may not extend to opposite face.	Lower density due to lack of material. Will appear round in shape. Color properties should vary due to wood material.
	Bark Pocket / Knot Hole	An area of bark where an encased knot has fallen out and only the bark surrounding the old knot remains. Often contains cracks filled with bark. Generally round. The hole does not continue to opposite face.	Lower density due to lack of material. Will appear round in shape. Color properties should vary due decayed wood material
Stain/Mineral Streak	Same as previous	An olive to greenish-black or brown discoloration of undetermined cause.	None
Clearwood	Same as previous	Earlywood and latewood growth transitions where no other feature is present.	None

Table 5.9. Classification results for red oak original feature classes compared with combined subclasses as percent classified correctly for each species and color space.

Color Space	Feature Groups	Feature Classification (% classified correctly)				
		Barkpocket	**Clearwood**	**Knot**	**Stain/Mineral**	**Overall**
RGB	Original	95	100	95	100	97.5
	Combined	93.55	100	86.21	84.21	91.0
HSI	Original	90	100	90	100	95.0
	Combined	90.32	100	82.76	89.47	90.6

feature classes. By combining the feature subclasses and re-evaluating the parameterrelationships using ANOVA, it is possible that a better set of classification variables could be identified to improve upon the accuracy.

5.4 Ability of Parameters to Differentiate Features Irrespective of Species

The forest products industry utilizes a variety of species in the manufacture of products. Any defect detection system would be required to identify defects in many different species; therefore, it is desirable to determine what parameters could be used to would provide good uniform classification between all species tested. It has been previously concluded that parameter differences do occur between species. Therefore, the relationships of parameters that were different between features, but similar between species, were used to develop a multi-species classifier. This classifier was compared to others developed by, forward and backward variable selection methods, and by selecting variables based on experience gained in the previous section. A discriminant classifier was developed separately for both the RGB, and HSI color parameters.

5.4.1 RGB Color Based Multi-species Classifier

The multi-species discriminant classifier developed using RGB color parameters is presented in Table 5.10. The parameters were selected based on the following relationships, which were discussed in detail in section 4.3. R_m was selected because of the number of significant differences between all features for all species. While B_m and G_m may be substituted for one another in hardwoods, G_m was chosen because of its contribution in the classification of white pine features. R_s was chosen because it presented a larger number of significant differences between feature types for the average for all species. ASP and X_m were chosen because they both were stated to be good for feature differentiation when used in a multiple parameter classifier. The classification accuracy for each feature type and the overall classification accuracies are listed in Table 5.11 for each colorspace and species. The discriminant classifiers developed based on the parameter relationships discussed provided higher classification accuracies than classifiers developed using all other possible combinations of parameters.

Table 5.10. Parameters selected for the RGB and HSI color based multi-species classifier for all species.

Base Color Parameter	**Selected Parameters**					
RGB	R_m	G_m	X_m	ASP	R_S	X_S
HSI	S_m	I_m	X_m	ASP	I_s	X_s

5.4.2 HSI Color Based Multi-species Classifier

A discriminant classifier was then developed for all species based the HSI color parameters, using the same methods applied to the RGB based classifier. The model parameters are presented in Table 5.10. The I_m was chosen because it has been shown to be the most important color parameter in all species. The I_m had more significant differences between features than any other color parameter. S_m was chosen because it proved to have more significant differences between features than the H_m for the three species. ASP, X_m, and X_s were included because of their contribution to classification accuracies, as discussed in section 5.3. The classification accuracies for all features and species are listed in Table 5.11. The discriminant classifiers developed based on the parameter relationships discussed provided equal or higher classification accuracies than classifiers developed using all other possible combinations of parameters.

5.4.3 Summary of Multi-species Classifiers

A multi-species classifier was developed using knowledge of parameter relationships of wood features, which will provide better classification accuracies than a classifier with all or randomly selected parameters. The relationships used in selecting classifier parameters were based on the significant differences between parameters measured from each feature type. The developed multi-species classifiers used two color mean parameters, one color standard variation parameter, one shape, and two density parameters indicating that these parameters should be included in defect classification methods to achieve the highest possible accuracies. While the multi-species classifier produced high classification accuracies, the individual species classifiers produced equal or higher classification accuracies indicating that individual species classifiers will provide greater accuracy than multi-species classifiers.

5.5.4 Comparison of RGB and HSI Color Spaces.

Most feature classification algorithms utilizing color information are based on the RGB colorspace, or a wavelength and intensity based colorspace such as, the HIS colorspace. Brunner et al. (1990) and Maristany et al. (1991) conclude that the

Table 5.11. Classification results presented as percent classified correctly for each species specific classifier and color space for multi-species classifiers.

Species	Color Space	Feature Classification (% classified correctly)				
		Barkpocket	Clearwood	Knot	Stain/Mineral	Overall
Red Oak	RGB	80	100	90	94.74	91.2
	HSI	90	100	80	100	92.5
Hard Maple	RGB	100	95	100	93.75	97.2
	HSI	85	95	95	93.75	92.2
White Pine	RGB	100	100	100	92	98.0
	HSI	88.89	95	100	92.3	94.0

appropriate color space for feature classification in wood should be determined for each particular application. Throughout this research, wood features have been described using parameters from both the RGB and HSI color space. In this section the ability of these two color spaces to differentiate between features is compared. Comparisons were made based on the classification results of the discriminant classifiers developed in sections 5.2 and 5.3, and on discriminant classifiers developed using color parameters only.

The results of the classification models developed for each species in section 5.3.2.1 and 5.3.2.2 for each color space is presented in Table 5.12. The overall classification accuracies were compared using McNemar chi-squared statisitc (Huberty, 1994). It was found that there were no significant differences between the overall abilities of the discriminant classifiers for any of the three species, indicating that the overall performances of the two color spaces were equal for each species. It was also discovered that there were no significant differences in classification accuracies for the two color spaces for feature types.

It was hypothesized that these results could be biased due to the inclusion of both shape and density parameters in the classifiers compared. Therefore, the ability of each color space to differentiate between species was then compared by developing models based on color parameters only. The color parameters were selected based on the relationships of feature parameters determined in Chapter 4 and proven effective in classification in section 5.3. For each species and color space classifiers were compared against all possible color parameter combinations to insure the best selection of parameters. In each case the classifier based on the relationships developed in Chapter 4 proved equal or better than any other developed classifiers. Comparisons were based on the number of correct classifications of features and not the overall accuracy of the classifier. Parameters selected for the classifiers are listed in Table 5.13. The results of these discriminant functions are listed in Table 5.14. None of the classifiers were significantly different in overall performance for any of the three species nor were there any significant differences in classification accuracies for each feature type.

Table 5.12 The classification results for RGB and HSI color based classifiers presented as percent classified correctly for each species and color space.

Species	**Color Space**	**Feature Classification (% classified correctly)**				
		Barkpocket	**Clearwood**	**Knot**	**Stain/Mineral**	**Overall**
Red Oak	RGB	95	100	95	100	97.5
	HSI	90	100	90	100	95.0
Hard Maple	RGB	100	95	100	94	97.3
	HSI	100	95	95	100	97.5
White Pine	RGB	100	100	100	92	98.0
	HSI	100	100	100	100	100.0

Table 5.13. Parameters selected for the comparison of color spaces.

Species	**Color Space**	**Selected Parameters**					
Red Oak	RGB	R_m	G_m	B_m	R_s	G_s	B_s
	HIS	H_m	I_m	I_s			
Hard Maple	RGB	R_m	G_{mrean}	R_s			
	HIS	H_m	S_m	I_m			
White Pine	RGB	R_m	B_m	R_s	B_s		
	HIS	S_m	I_m	I_s			

Table 5.14. Classification results between color spaces using color parameters only.

Species	**Color Space**	**Feature Classification (% classified correctly)**				
		Barkpocket	**Clearwood**	**Knot**	**Stain/Mineral**	**Overall**
Red Oak	RGB	100	95	90	84.2	92.3
	HSI	90	100	85	78.6	88.4
Hard Maple	RGB	100	95	100	93	97.0
	HSI	92.8	95	100	100	97.0
White Pine	RGB	100	100	95	84.2	94.9
	HSI	100	100	95	76.9	93.0

It can be concluded that there are no significant differences between the two colorspaces when used for classification of the features and species examined in this research. The similarity of the two colorspaces in classification of features can be explained in part by the identical trend between feature differences determined for the R_m, G_m, B_m, and I_m parameters. It can be concluded from these results that for the classification of ingrown knots, bark pockets, stain, and clearwood there is no significant difference in discriminant classification using either RGB or HSI color parameters for red oak, hard maple, and white pine.

5.6 Chapter Summary

In this Chapter discriminant classifiers were developed using forward, backward, all possible combinations, and knowledge determined in Chapter 4. The classification accuracies of these discriminant classifiers were compared to demonstrate that an in-depth knowledge of how parameters are related between features can be used to develop the best possible classification functions for individual species and multi-species. The methods for variable selection used in Chapter 4 have been verified by comparing the classification accuracies of the various classification functions. It was determined that the required classification parameters do vary between species. It was found that the RGB and HSI color space parameters provide the same classification accuracies for the species and features studied. When differentiating between difficult to classify features and species, color parameters were shown to improve classification accuracies. As suggested by Conners et al. (1985), Maristany et al. (1991), two color parameters are required for the optimal classification accuracy of features. While density and shape parameters prove poor classifiers when used alone, combined with other parameters they both increase the performance of classifiers regardless of the species or colorspace. Although the conclusions are based on manually selected feature regions from homogenous feature classes, the methods used to develop parameter relationships in this research can be used to increase the performance of classification methods for other features, species, and parameters.