Colour image analysis for classification of dried fruit

Karen Henry and Gerhard de Jager

30 May 1996

Abstract

A computer vision system that would analyse and classify dried fruit, in particular dried peaches, pears and apricots, was investigated. Each class (grade) of dried fruit was described by a normal probability distribution in a chosen colour feature space. In order to colour classify dried peaches the dark coloured core area was isolated from the peach using a Morphological Top-Hat transform. Because certain of the dried pear grades were graded according to blemish size, a blemish thresholding technique, based upon the information in the colour band histograms was developed. The colour classification system and the blemish isolation process produced good test results.

1 Introduction

South Africa has a very successful and lucrative fruit industry. The quality control of both fresh and dried fruit is very important. The quality grading of dried fruit is done either manually or by automation. Existing machine vision systems that classify fruit according to their flesh colour are very successful in grading fresh fruit, but as found by South African Dried Fruit, are only partially effective in sorting dried fruit. At S.A. Dried Fruit only dried apricots are machine vision sorted; dried pears and peaches must be sorted manually because of the dismal machine vision sorting results. In human visual inspection of dried fruit, the colour of the fruit is often the main determinant of its grade; in specific cases the presence of blemishes and geometrical fault are also incorporated in order to determine the fruit grade. A machine vision system was required that would cater for the following :

1. The majority of dried fruit are allocated to a grade according to their flesh colour. A colour classification system was required that could differentiate between the image colours of different grades of dried fruit. This colour classification scheme would be the basis of a final dried fruit classification system that would tabulate and grade each dried fruit independently, into predefined classes, according to its colour.

2. In the South African dried fruit industry, main dried fruit types include dried apricots, peaches and pears. Before a peach is sulphur dried, the pip of the fruit is removed, leaving a core area that is much darker in colour than the remainder of the peach

flesh. If a peach was automatically graded according to its entire flesh colour, it would be graded as REJECT, because the core area is the colour of a peach blemish. The solution would require that any dried peach image be divided into two parts; the dark coloured core area would have to be isolated from the remaining lighter coloured peach flesh. A process would have to be found that would accurately segment the core area from a dried peach within an acceptable time frame. This segmentation process would have to be performed on every single dried peach that was examined. The remaining non-core area of the flesh would be used to classify the peach according to its colour.

3. The grade specifications (i.e. the rules according to which a fruit is graded), although largely dependent on colour, include examination of two further areas; the fruit geometry, and in the specific case of dried pears, blemish area or size. Certain grades of dried pear were graded entirely according to the existence and size of those blemishes found on the flesh. Classification of these grades of dried pear require the recognition, isolation and sizing of black blemishes upon the dried pear. This procedure needed to be automated. Both the dried fruit geometry and the blemish size are currently included in the final manual grading decision of dried fruit. The necessity of automatic geometry validation needed to be investigated.

The aim of this investigation was to design an accurate, efficient and time-effective machine vision grading system for dried fruit. The proposed method consists of a number of steps. Because colour is a major factor in the grading of dried fruit, a colour co-ordinate system which would accurately differentiate between the colours of dried fruit, was investigated. Segmentation techniques were required to segment (1) the fruit from the image background, (2) black blemishes from within dried pears, and (3) the dark core area of a dried peach from the peach flesh. For the first two problems, thresholding techniques based upon the colour histograms of the image proved very successful. In order to segment the peach into core and non-core areas, the geometry of the peach image was considered; this lead to an investigation of the theory and operations of Morphology. The Top-Hat Transform in particular, was very successful in isolating the core area of a dried peach. Finally the colour features were utilised to classify each dried fruit through the use of a supervised classifier based upon the Quadratic Discriminant Classification Rule.

Section 2 introduces several colour models, including the one selected for dried fruit classification. In Section 3 the Top-Hat transform as a segmentation technique, is detailed, as well as the histogram thresholding technique and its application to the segmentation problems of object(dried fruit) and blemish isolation. Section 4 details the basis of the Quadratic Discriminant Classification Rule, and its statistical relevance to the chosen feature space. Section 5 and 6 give experimental results and conclusions.

2 Background

2.1 Colour Models

In the selection of a colour system, two issues were important :

1. The colour system should clearly differentiate between the colours of different grades of dried fruit.

2. The overall processing time of any commercial machine vision system is very important. Any colour system that enhanced the overall processing speed of the classification scheme would be considered seriously; conversely any colour system that lengthened this process, would have to be viewed with caution.

Of those machine vision systems that have been developed to analyse and classify fruit according to their colour, the fruit sorter designed by Kay [4] and implemented by Teklogic, the KROMCO fruit sorter, and a design for a fruit sorter by Guedalia [5] all make use of the HSI colour model, the KROMA SORTER makes use of the RGB colour model.

Ohta *et al* [15] documented an investigation of colour segmentation, comparing a broad range of colour models. They conducted a very thorough and comprehensive investigation of more than 100 colour features (including RGB, YIQ, HSI, normalised RGB, and La*b* systems), in order to find those colour features which achieved a colour segmentation comparable to that of the colour features X_1,X_2,X_3 obtained by the Karhunen-Loeve transformation [14]. Three colour systems performed particularly well: the La*b*, the HSI, and the derived I_1,I_2,I_3 ($I_1 = \frac{R+G+B}{3}$, $I_2 = \frac{R-B}{2}$, $I_3 = \frac{2G-R-B}{4}$) colour systems.

Only the $L^*u^*v^*$ and the $L^*a^*b^*$ models, as described by [7] are recommended by the Commission Internationale de l'Eclairage (CIE) for uniform colour difference evaluation. It was found, however, that the La*b* model, although producing good results was too costly; from the computational point of view, the non-linear transformations incur far more cost than linear transformations. An investigation of both the RGB and the HSI colour systems was therefore done. Each of the variables in the two colour spaces were analysed independently to ensure that redundant features were discarded. In the RGB space, the Blue band values for each type of dried fruit were either zero, or closely replicated the Green band values; for this reason only the Red and Green features were used within the RGB feature space. In the HSI feature space, the Hue and Intensity values provided more information on colour changes than did the Saturation feature. Using an assortment of dried fruit types, the separation and deviation of the H,I and R,G feature variables within *good* and *bad* grades were examined. The Hue values were normalised to the range $(0, 255)$. For each of three fruit types the average value for each grade (average over three fruit samples), the Euclidean distance d_E (see Section 4) between the two grades of good and bad fruit, and the related standard deviation value are given in Table 1. For each dried fruit type the R_{avg}, G_{avg} values of good and bad samples were well separated; the H_{avg}, I_{avg} values were not well separated for dried pears. The R_σ, G_σ for good dried pears and apricots was very small, whereas the H_σ, I_σ values for dried peaches were small. As expected, in all cases the bad samples had larger standard deviations than good samples, but none were extremely large. From these standard deviations, well clustered data classes about the class average values were expected, particularly for the higher quality grades of fruit. For both speed of processing and good feature space clustering, the R, G features were adopted to describe dried fruit colour. For each fruit the Red band pixel values would be averaged to obtain R, and similarly the Green band values would be averaged to obtain G. Scatterplots in an (R,G) colour coordinate system for good and bad fruit types were checked to ensure reasonable class separabilities between (R,G) feature vectors in this feature space; an example is shown in Figure 1.

TABLE 1 : Average values and Euclidean distances(d_E) in R,G or H,I feature space								
		Red and Green values				Hue and Intensity values		
FRUIT	STATS	GOOD	d_E	BAD	STATS	GOOD	d_E	BAD
PEAR	R_{avg}	205	81	124	H_{avg}	18	2	20
(DRIED)	R_σ	10	20	30	H_σ	11	18	29
	G_{avg}	123	78	45	I_{avg}	211	11	200
	G_σ	5	10	15	I_σ	15	8	23
APRICOT	R_{avg}	180	80	100	H_{avg}	50	65	240
(DRIED)	R_σ	5	45	50	H_σ	16	19	35
	G_{avg}	60	30	30	I_{avg}	180	80	100
	G_σ	5	7	12	I_σ	4	46	50
PEACH	R_{avg}	170	50	120	H_{avg}	15	25	245
(DRIED)	R_σ	22	16	38	H_σ	5	3	8
	G_{avg}	60	24	36	I_{avg}	170	40	130
	G_σ	36	14	50	I_σ	12	28	40

Table 1: Average values for dried fruit in the colour bands: H,I,R,G

3 Method

3.1 Pre-processing

Segmentation is a process of classifying the set of given image pixels into subsets according to their distinguishing characteristics [12]. The dried fruit must be segmented from the image background, so that the dried fruit can be processed according to its own unique characteristics. The colour white was found to have high values of red, green, and blue, whereas dried fruit have very low, often zero, blue values. Jain *et al* [2] used this knowledge (of high R,G,B values in white) to successfully threshold white backed address labels from magazine covers. Dried fruit upon a white background were found to produce well separated modal peaks in the blue band histogram. The blue band therefore was used to threshold dried fruit from the image background using the blue band histogram. The Discriminant Analysis thresholding method [9] produced the threshold value closest to the visual, or manual threshold value. In Figure 2 the success of the Discriminant Analysis method in removing the dried fruit from the background, is seen.

3.2 Segmentation using a Morphological Top-Hat Transform

In grey scale images, if you subtract an *opened* [3] image from the original image, you will be left with features such as peaks and points of high curvature. In Morphology, the operation of subtracting an *opened* image from the original image has been labelled the **Top-Hat**

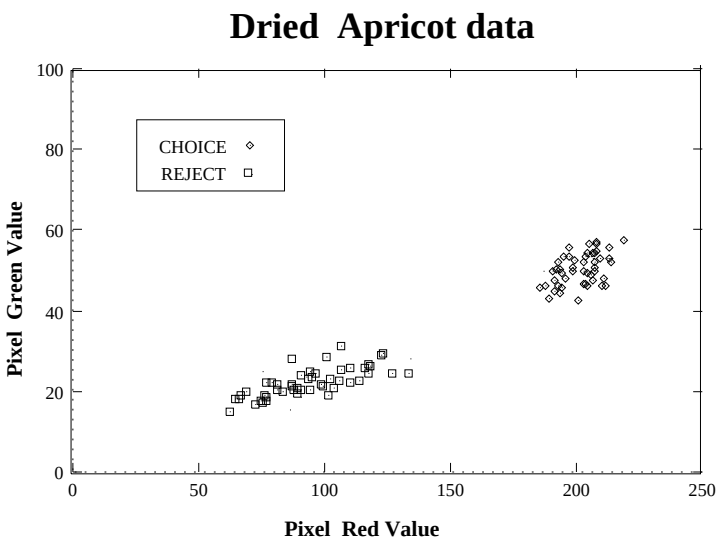

Figure 1: Scatterplot of the Red and Green band pixel values of a *good* and a *bad* dried apricot

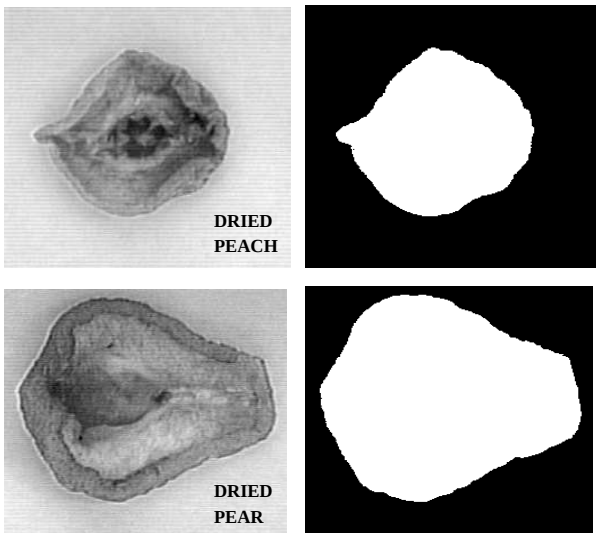

Figure 2: Thresholding of the background of two dried fruit

Transform and is defined by :

$$\mathrm{HAT}(f) = f - (f \odot g)$$

where g is the chosen structuring element, and $\odot$ is the opening operator [3]. In the case of a peach, the manual removal of the pip results in a 'core area' which is a much darker red, brown than the remainder of the peach flesh. The Top-Hat Transform was found to be very successful in the segmentation of the core region from any given dried peach. The Red band pixel values of the peach closely follow the rise and fall in the corresponding Intensity ($I = \frac{R+G+B}{3}$) pixel values, and consequently accentuate the geometrical 'indentation' or valley of the core region. This close correlation between Red band and Intensity band image values is attributable to the consistently low blue (approximately zero) and green band values in dried peaches. The Top-Hat transform operator was therefore applied to the Red band values of the RGB image. The central region of the peach where the core was found, was isolated using the centre-of-mass algorithm by Levine [8]. Only a square region about that centre-of-mass was then processed. An effective structure element had to be selected for the Top-Hat transform; in order to highlight the 'valley' of the core, the structure element had to have a diameter larger than the diameter of the core (valley) it was attempting to isolate. A disk element was the most accurate and isolated the core area in the shortest time. The radius of the disk was dependent upon the size of the image core pixel area. The original images consisted of four peaches; each peach occupied a 240×240 pixel area. A window area of 120×120 about the centre-of-mass was processed using the Top-Hat transform. The process remained completely successful but was much faster if the 120×120 image was reduced in size by a magnitude of 8 to a 15×15 pixel area. A disk shaped structure-element of radius 5 pixels was found to be the most successful on these 15×15 pixel areas. The result of the Top-Hat transform on two different dried peaches is shown in Figure 3. Once the core had been identified by the transform, a grey level threshold was used to generate a mask for the core area. Five thresholding methods were tested: Pun's Entropy Method [11], Tsai's Moment-Preserving Method [13], Kittler's Minimum-Error Method [6], Brink's Maximum Correlation Method [1] and Otsu's Discriminant Analysis Method [9]. The Discriminant Analysis automatic thresholding algorithm was found to be the most successful.

3.3 Identifying and Isolating Blemishes in Dried Pears

Dried pears are graded into four classes by South African Dried Fruit (SADF). The 3rd and 4th classes consist of those dried pears that have black blemishes. If blemishes upon pears could be automatically isolated and measured, allocation of these two grades of pear would be simple. Each of the three bands of the RGB image were investigated in order to find that band which, when thresholded, achieved the best blemish isolation. The Red and Green colour bands, when individually thresholded and the resulting masks combined, produced the best isolation of black blemishes. The two thresholding procedures produced binary images i and j. The binary images i and j were then combined as follows to produce a binary image k :

$$k = i + j \qquad \text{where } + \text{ is a LOGICAL OR procedure}$$

Again, visual examination of the best manual thresholding value was used to discover the most appropriate automatic thresholding technique for the task. The Entropy [11] automatic

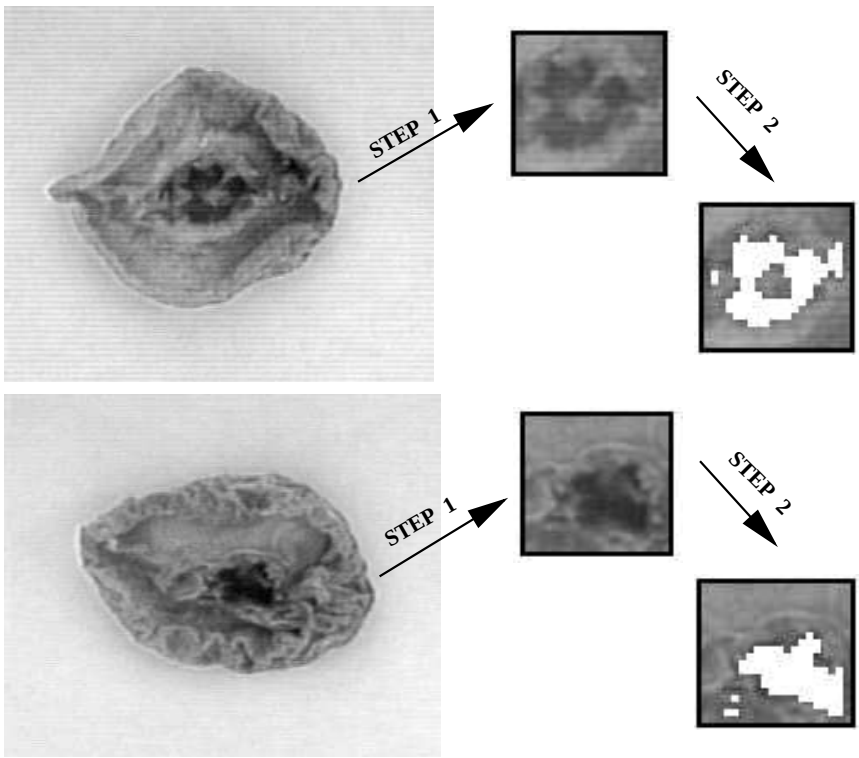

Figure 3: The results of the Top-Hat Transform on the image Red band values

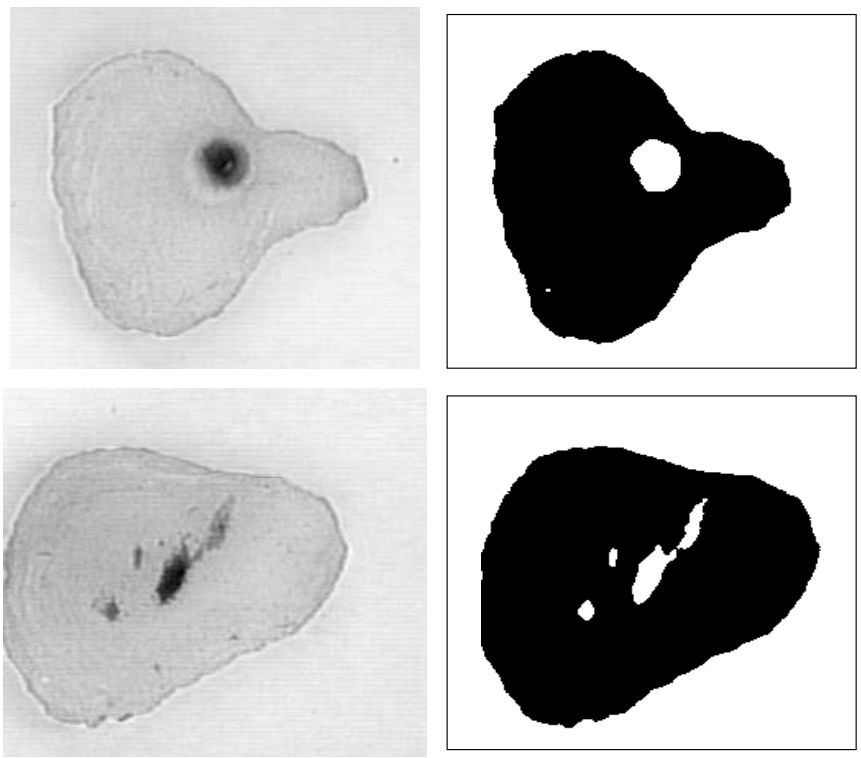

Figure 4: Thresholding of black blemishes from dried pears

thresholding technique, which does not assume a bimodal distribution, was found to be the most successful in isolating blemishes on a dried pear.

The existence of blemishes, as well as their pixel area (or diameter), is available to be used in a dried fruit classifier to aid the colour classification scheme. The above technique succeeded in isolating every single black blemish on each of the 42 dried pears (from both class 3 and 4) that were tested. Two examples are represented in Figure 4.

4 Colour-based multivariate Quadratic Discriminant classification

4.1 Statistical versus Euclidean multivariate data

In the dried fruit grading system, the R and G variables constitute the multivariate data of the system, where each dried fruit has its own average R,G values that describe its individual colour information. Much of the information contained in the multivariate data can be assessed by calculating and using certain summary numbers from the multivariate data sets; these numbers are known as *descriptive statistics* and involve the concepts of **sample mean**, the **sample variance** and the **sample covariance** [10].

In general the *Euclidean distance* from a point $P = (x_1, x_2, ..., x_p)$ to a point $Q = (y_1, y_2, ..., y_p)$ is :

$$\begin{aligned} d(P,Q) &= \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \cdots + (x_p - y_p)^2} \\ &= \sqrt{(\mathbf{x} - \mathbf{y})^T (\mathbf{x} - \mathbf{y})} \end{aligned} \quad (1)$$

where T implies the transpose operator on the preceding vector or matrix. In *Euclidean distance* measurement, each coordinate contributes equally to the calculation of the distance; this balance is not always satisfactory for statistical purposes. In some cases the variable measurements are subject to different rather than similar extents of fluctuations. What this means is that two variables X and Y may not undergo similar variations, as is assumed in the *Euclidean distance* case; changes in the X variable measurements, for example, may tend to be obviously greater than changes in the Y variable. In these cases it would be preferable to weight those coordinate variables such as X, that fluctuate greatly, less heavily than variables like Y. These distance measures that account for differences in variable variation are referred to as *statistical distances*. For variable measurements which **do not vary independently**, the *statistical distance* measure between points P and Q is :

$$d_s(P,Q) = \sqrt{(\mathbf{x} - \mathbf{y})^T \mathbf{A} (\mathbf{x} - \mathbf{y})} \quad (2)$$

Ordinarily $\mathbf{A} = \mathbf{S}^{-1}$ where the sample covariance matrix $\mathbf{S}$ contains the sample variances and covariances [10]

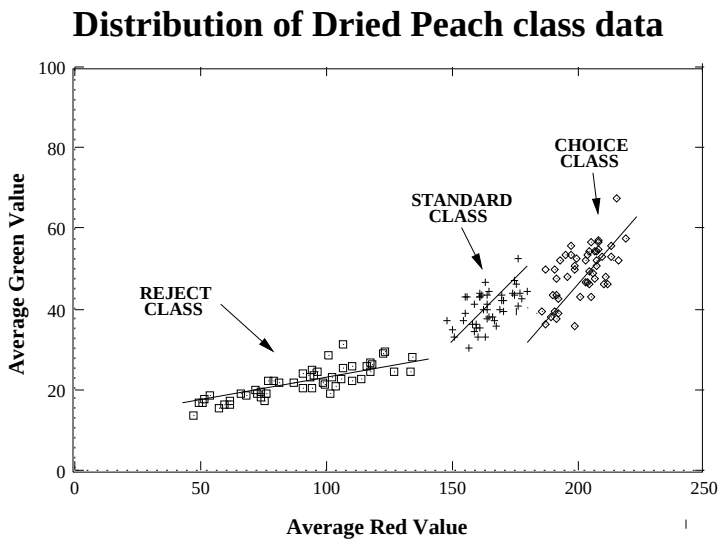

Figure 5: A scatterplot showing class differences in position and in relative variations within X and Y

Figure 5 is a scatterplot which describes three classes of dried peach: Choice, Standard and Reject, in a 2-dimensional R,G feature space. Each point in the scatterplot describes the average R,G values for one individual dried peach (with the core area removed). In each class, the data varies more in distance along one axis than the other; the data are elliptical in

nature. For this reason the *statistical distance* $d_s(P,Q)$ of Equation 2 would serve as a more suitable distance measure technique than the Euclidean measure.

This distance representation of the data distribution facilitates the use of a probability density function to describe each dried fruit class distribution. Gaussian or normal probability density functions are based on a statistical square distance measure of the samples. If a normal probability distribution is used to model each of the dried fruit data sets, then for each class of fruit w_i, a data-based normal distribution $p_i(\mathbf{x})$ must be generated in the R, G feature space :

$$p_i(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{2}{2}} \ |\mathbf{C_i}|^{\frac{1}{2}}} \ e^a$$

$$\text{where} \quad a = [-\frac{1}{2} \ (\mathbf{x} - \mathbf{m_i})^T \ \mathbf{C_i}^{-1}(\mathbf{x} - \mathbf{m_i})]$$

For this 2 band problem, $\mathbf{m_i}$ is a 2-dimensional vector containing both the **R sample mean** and the **G sample mean** and $\mathbf{C_i}$ is a (2 x 2) **sample covariance matrix**. If each dried fruit class was defined by a normal distribution, which by definition has no finite borders, any dried fruit observation would have a probability of being in all of the four classes - albeit an extremely small probability of occurrence in some classes. The sum of the probability densities multiplied by the corresponding class probabilities defines the combined probability surface

$$f(\mathbf{x}) = \sum_{i=1}^{M} p(\mathbf{x}|w_i).p(w_i)$$

Because the covariance matrices of each density function are unequal, the discriminant or decision boundary separating any two class distributions is quadratic. The parameters of class mean vector and covariance matrix of the normal probability distribution for each class, would have to be determined in a training run.

4.2 Optimal Classification Rules

A good classification procedure should result in few misclassifications. Over and above the basic discriminant rule, there are two additional features that an 'optimal' classification rule should possess [10]: one feature takes the *a priori* class probabilities $p(w_i)$ and the probability of occurrence into account; the other takes the *cost of misclassification* into account. The *costs of misclassification* can be defined by a cost matrix such as the following example :

True Population	Classify as: Choice	Reject
Choice	0	c(R | C)
Reject	c(C | R)	0

The costs are :

1. Zero for correct classification

2. c(C | R) when an observation from Reject is incorrectly classified as Choice

3. c(R | C) when a Choice observation is incorrectly classified as Reject.

Johnson [10] suggests the following classification rule for assigning $\mathbf{x}$ to one of classes R1 and R2 :

$$\begin{aligned} R1 \ : \ \frac{f_1(\mathbf{x})}{f_2(\mathbf{x})} &\geq [\frac{c(1|2)}{c(2|1)}][\frac{p_2}{p_1}] \\ \text{[Density ratio]} &\geq \text{[cost ratio] [prior probability ratio]} \\ R2 \ : \ \frac{f_1(\mathbf{x})}{f_2(\mathbf{x})} &< [\frac{c(1|2)}{c(2|1)}][\frac{p_2}{p_1}] \\ \text{[Density ratio]} &< \text{[cost ratio] [prior probability ratio]} \end{aligned} \tag{3}$$

where $f_1(\mathbf{x})$, $f_2(\mathbf{x})$ are the class density distributions, and p_1, p_2 are the probability of occurrences.

The *a priori* probabilities for each class of dried fruit are taken to be equal; what this means is that a dried fruit $\mathbf{x}$ has the same probability of being in Grade 1, as it has for any of the other classes, which is an acceptable assumption under specified test set conditions. The costs of misclassification would need to be determined by the respective company, however, for test purposes of this dried fruit classifier, costs of misclassification have been assumed equal. Johnsons [10] classification rule of Equation 3, under the above two assumptions becomes :

$$\begin{aligned} R1 \ : \ \frac{f_1(\mathbf{x})}{f_2(\mathbf{x})} &\geq 1 \\ R2 \ : \ \frac{f_1(\mathbf{x})}{f_2(\mathbf{x})} &< 1 \end{aligned} \tag{4}$$

for a 2-class classifier. This means that a pattern $\mathbf{x}$ has a degree of membership in every class, but it is a member of the class to which it has the highest degree of membership, that is

$$\mathbf{x} \ \epsilon \text{ class } i \qquad \text{if } \ p_i(\mathbf{x}) > p_j(\mathbf{x}) \qquad \text{for } j \neq i$$

This section has described how each class of dried fruit is represented by a normal density distribution. Because the covariance matrices of each distribution are unequal, the distributions form quadratic discriminant boundaries between the classes. Any test sample of dried fruit would be classified according to these boundaries, and according to the classification rules of Equation 4.

4.3 Rule Induction from Training Sets

Sample groups or *labeled sets* of dried fruit were hand selected by trained SADF staff. For three fruit types: dried peach, dried apricot, and dried pear, 204, 204 and 128 sample units were selected respectively. The full data set or sample group of each fruit type was made up of four classes: Choice, Standard, Sub-Standard and Reject. The *hold-half-out method* [10] was used to evaluate the classifier. In the hold-half-out method the total labeled set is divided in half; one half is used to construct the decision rule or rather is used to set up the parameters

of the normal probability distributions for each class, and the other half is used to test the accuracy of the decision rule. The roles of each set are then switched and the test repeated. Once each half has played the role of both 'training set' and 'test set', the error rates for both 'test sets' are combined to establish the final error estimate. This method is presumed to exaggerate the error rate of the classifier. 26 Dried apricots or peaches, and only 16 pears were used to define the parameters of each colour class. These training sets should be much larger; the larger the training set, the more thoroughly the colour of a class is defined, and consequently fewer misallocations will occur.

5 Experimental Results and Discussion

The procedure used for all of the results is the *Quadratic Discrimination Classification Rule*(QDCR) of Equation 4. There were two RUNS or sets of results for each fruit type. RUN A tests included those grades that were specified according to colour only. RUN A was meant only to examine the success of the colour classification scheme. RUN B tests, on the other hand, were performed on all fruit classes. Sub-Standard dried apricot and peach classes consisted essentially of those fruit that have geometrical faults (i.e. too small, squashed, or blemished); Sub-Standard and Reject class pears were defined by the presence and size of one or more black blemishes. RUN B highlighted whether or not the grading of fruit could be done according to colour only. The same training sets were used in each of the runs. Approximately half of the selected set were images of the front of the dried fruit, the other half of the back (flat surface) of the fried fruit.

The actual test results for dried peaches are presented. The test results for dried apricots and dried pears are listed. Before each peach was classified, the core area was isolated using the Top-Hat transform. The results for each RUN were divided into the different classes of Choice(C), Standard(S), Sub-Standard(SS) and Reject(B). The number of dried peaches in the manually selected test set, is specified under 'Manually Allocated'. The class to which the Colour Classifier allocated each peach in this set is listed under 'Classifier Allocation'; the corresponding percentage of the total *test set* is listed under '% of Test Group'. The feature vectors for these particular classes of dried fruit have been plotted in (R,G) feature space in Figures 6, 7, and 8. The ellipses in these diagrams are present merely to indicate the grouping of the vectors in each class.

5.1 Discussion

RUN A tests, which included only those grades of dried fruit where flesh colour was the deciding factor, produced the following results:

- For dried peaches: a minimum allocation success of 84%, and an average success of 87% over 3 classes
- For dried apricots: a minimum allocation success of 96%, and an average success of 97% over 3 classes

RUN A : Dried Peaches				
Manually Allocated		Classifier Allocation		% of Test
Number	Class	Number	Class	Group
51	C	44	C	86%
		7	S	14%
		0	B	0%
51	S	43	S	84%
		6	C	12%
		2	B	4%
51	B	47	B	92%
		0	C	0%
		4	S	8%

Table 2: Class allocations produced by the QDCR Classifier for those classes of dried peaches that were grouped by flesh colour(hue) alone

RUN B : Dried Peaches				
Manually Allocated		Classifier Allocation		% of Test
Number	Class	Number	Class	Group
51	C	44	C	86%
		7	S	14%
		0	SS	0%
		0	B	0%
51	S	37	S	72%
		6	C	12%
		8	SS	16%
		0	B	0%
51	SS	22	SS	42%
		0	C	0%
		9	S	18%
		20	B	40%
51	B	41	B	80%
		0	C	0%
		2	S	4%
		8	SS	16%

Table 3: Class allocations produced by the QDCR Classifier for all classes of dried peaches, including those that were allocated according to flesh colour, fruit size, blemish presence, and other geometrical flaws

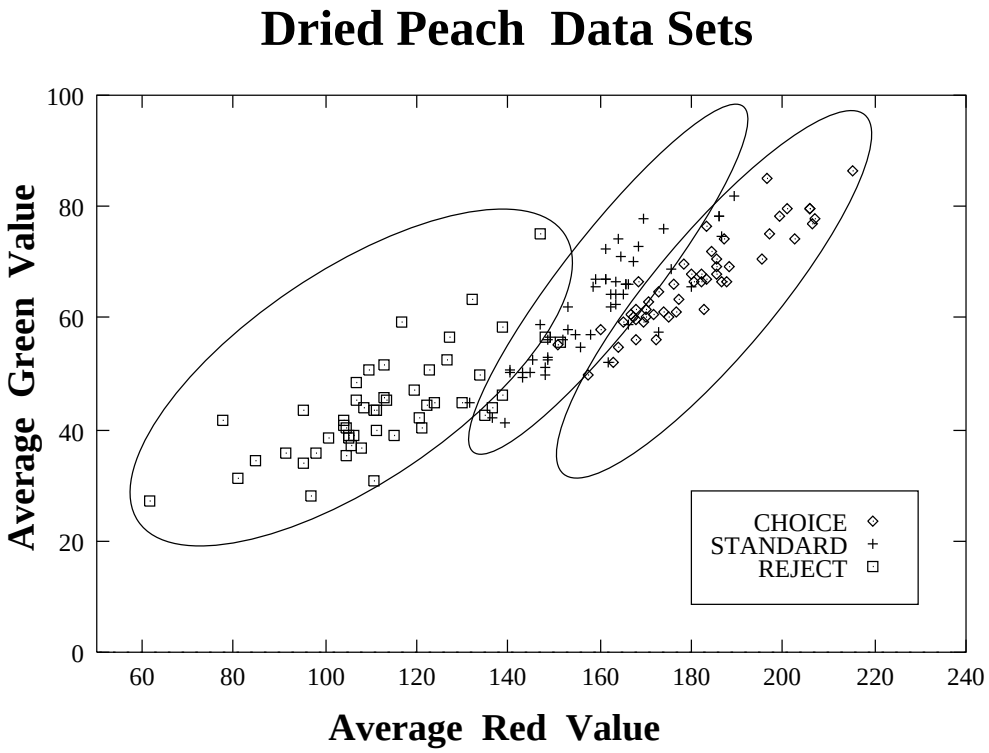

Figure 6: Feature vector distribution of 3 dried peach classes

- For dried pears: a minimum allocation success of 91%, and an average success of 93% over 2 classes.

RUN B tests, which included all grades of dried fruit, produced the following results:

- For dried peaches: a minimum allocation success of 42%, and an average success of 70% over 4 classes
- For dried apricots: a minimum allocation success of 12%, and an average success of 68% over 4 classes
- For dried pears: a minimum allocation success of 40%, and an average success of 49% over 4 classes.

The hold-half-out evaluation method is considered pessimistic, therefore better results than those above may be expected. RUN B tests produced various levels of success, but were particularly unsuccessful for dried pears. Because the Sub-Standard classes of dried peach and apricot are largely the geometrically flawed fruit, the SS class colour distribution encroaches into all classes. In order to be successful, it is recommended that the geometry of the dried fruit be analysed in a pre-processing run, before colour classification occurs. In the case of dried pears, Sub-Standard and Reject class pears are defined by the presence and size of one or more black blemishes. The remainder of the pear flesh, in these two classes, would often

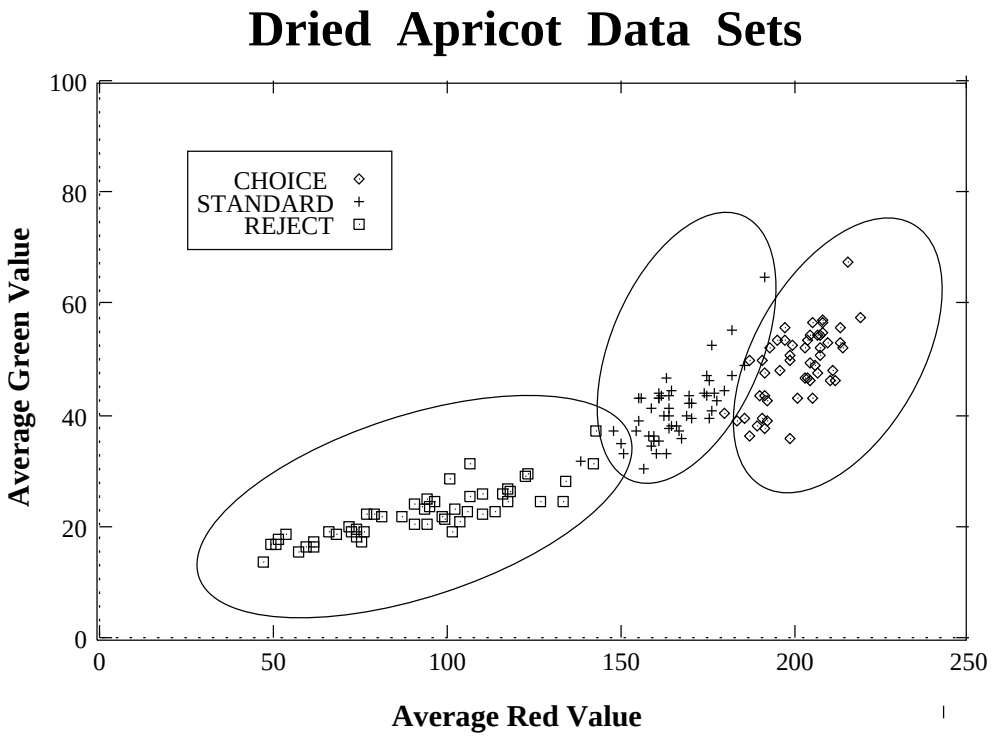

Figure 7: Feature vector distribution of 3 dried apricot classes

be described as Choice grade in colour. The width of the black blemish(es) alone is used to divide the pears into Sub-Standard and Reject classes. The remainder of the pear flesh, in these two classes, would often be described as Choice grade in colour. To ensure successful classification of dried pears, the blemish areas (and width) must be processed and considered independently. In order that successful dried pear classification occurs, it is recommended that an initial blemish isolation and blemish sizing process is performed on each dried pear. Depending upon the blemish size, the pear would be allocated to either the Sub-Standard or Reject class. Those dried pears that were found to be blemish-free, would be submitted to the Colour classification process for classification into the Choice, Standard or Sub-Standard classes.

6 Conclusion

In this paper a system of analysing and classifying dried fruit is proposed. The proposed classification method relies largely upon a colour classification discriminant rule, but requires both pre-processing of the image, as well as additional processing features depending upon the unique characteristics of the dried fruit and the grading specifications. The pre-processing involves the isolation of the fruit within the image using colour band histogram thresholding. The additional features necessary to support the colour classification rule are the following. The dried peach core areas must be removed using the morphological Top-Hat transform.

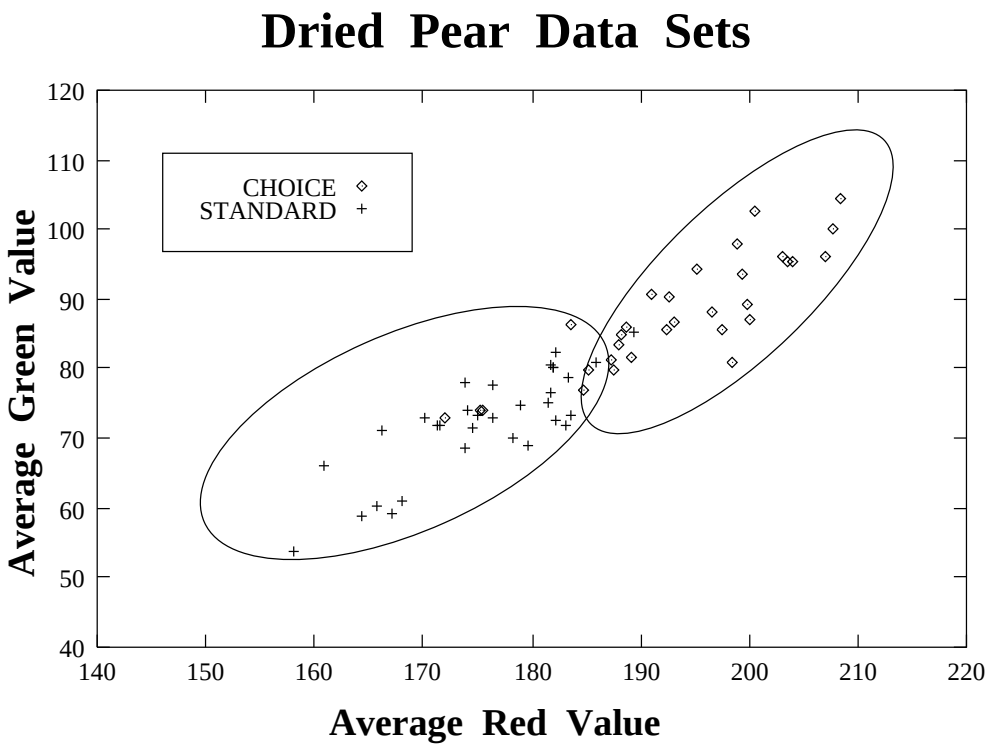

Figure 8: Feature vector distribution for 2 classes of dried pear

The geometry of dried apricots and peaches should be analysed so that those fruit belonging to the Sub-Standard classes can be isolated. Black blemishes upon dried pears must be isolated using colour band histogram thresholding, and the blemish size used to aid the final classification decision of the colour classifier. The proposed techniques produced very good results and if implemented should have acceptable processing speeds because the techniques were selected for their simplicity as well as their accuracy.

References

[1] Brink A.D. The selection and evaluation of grey-level thresholds applied to digital images. Master's thesis, Rhodes University, South Africa, 1987.

[2] Jain A.K. and Chen Y. Address block location using color and texture analysis. *CVGIP Image Understanding*, 60:179–190, September 1994.

[3] Dougherty E.R. *An Introduction to Morphological Image Processing*, volume TT9 of *SPIE Tutorial Texts in Optical Engineering*. SPIE Optical Engineering Press, 1992.

[4] Kay G.R. Colour analysis and the classification of fruit. Master's thesis, University of Cape Town, Cape Town, South Africa, 1992.

[5] Guedalia I.D., Edan Y., and Pasternak H. Multi-sensor quality classification of tomatoes. ASAE Paper No. 94-6032, ASAE, 1994.

[6] Kittler J. and Illingworth J. Minimum error thresholding. *Pattern Recognition*, 19:41–47, 1986.

[7] Tajima J. Uniform color scale applications to computer graphics. *Computer Vision, Graphics, and Image Processing*, 21:305–325, 1983.

[8] Levine M.D. Feature extraction: A survey. *Proceedings of the IEEE*, 57:1391–1405, August 1969.

[9] Otsu N. A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-9:62–66, 1979.

[10] Johnson R.A. and Wichern D.W. *Applied Multivariate Statistical Analysis*. Prentice-Hall Series in Statistics. Prentice-Hall, Inc., 1982.

[11] Pun T. A new method for gray-level picture thresholding using the entropy of the histogram. *Signal Processing*, 2:223–237, 1980.

[12] Niblack W. *An Introduction to Digital Image Processing*. Prentice/Hall, 1986.

[13] Tsai W. Moment-preserving thresholding: A new approach. *Computer Vision, Graphics, and Image Processing*, 29:377–393, 1985.

[14] Pratt W.K. *Digital image Processing*. John Wiley and Sons, Inc., second edition, 1991.

[15] Ohta Y., Kanade T., and Sakai T. Color information for region segmentation. *Computer Graphics and Image Processing*, 13:222–241, 1980.