Statistics and Applications
Volume 9, Nos. 1&2, 2011 (New Series), pp. 15-29

Min-max representation of features for grading cured tobacco leaves

D S Guru[1], P B Mallikarjuna[1], S Manjunath[1] and M M Shenoi[2]
[1]*Department of Studies in Computer Science, Manasagangothri, University of Mysore, Mysore-570006, Karnataka, India*
[2]*Central Tobacco Research Institute, Research station, Hunsur-571105, Mysore District, Karnataka State, India*

Abstract

In this paper, a new model to grade cured tobacco leaves using symbolic data is proposed. Color features such as average and standard deviation of hue of tobacco leaves are extracted using the Munsell color system. The extracted features of each grade are aggregated and represented in the form of inter-valued type data and stored in the knowledgebase. For a given test sample of a cured tobacco leaf, the features which are of crisp type are extracted and compared with the representatives stored in the knowledgebase. A symbolic classifier has been used for comparison purpose. In order to corroborate the efficacy of the proposed model, experiment has been conducted on real dataset consisting of 887 sample images of 12 different grades of cured tobacco leaves. The effect of grading accuracy under varying size of database has been studied and achieved an average grading accuracy of about 91%. A qualitative comparative analysis of the proposed model with other state of the art models is presented to study the elegance of the proposed model.

Key words: Cured tobacco leaf; Symbolic representation; Symbolic classifier; Grading

1 Introduction

Tobacco is a commercial crop in many countries like China, India, Brazil, United States, European Union, Zimbabwe, Indonesia, Malawi and Russia because of its high economic value. In India, Flue Cured Virginia (FCV) tobacco crop is cultivated in Andhra Pradesh and Karnataka states (Manickavasagam et al., 2007). The southern zone of Karnataka produces quality filler tobacco of export potentially. Nearly 80% of the tobacco produced from Mysore and Hassan districts of Karnataka state is exported annually to several countries to meet the demand of multinational cigarette manufacturing companies. Nearly seventy thousand farmers are cultivating this crop in this belt as it is a life line for them. In fact, tobacco crop created a gainful employment to several lakhs of people in this belt.

Farmers will do grading based on the quality of the flue-cured tobacco leaves before taking them into a market. Quality inspection of the flue-cured tobacco leaves plays a crucial role in quality assurance, since the quality of the flue-cured tobacco leaves determines the quality of tobacco products. According to Indian Council of Agricultural Research (ICAR) cured tobacco leaves are graded based on the position they grow on the stalk, color and quality. Quality is decided based on extent of damage due to blemish and spots. Based on the position they are divided into four large groups viz., primings (P), lugs and cutters (X), leaf (L) and tips (T) [1]. These large groups are divided into sub groups based on colors – lemon (L), orange (O), J-style (J), mahogany (R) and green (G). The groups are formed by combining position and color categories. For example consider lemon color, the sub groups formed are primings lemon (PL), lugs and cutters lemon (XL), leaf lemon (LL) and tips lemon (TL). Similarly sub groups are formed with other colors also. Again these sub groups are divided into 4 grades based on quality. For example, 4 grades are available in XL subgroup viz., X1L, X2L, X3L and X4L . Similarly all other subgroups have 4 grades.

A few attempts have been made especially on grading of flue-cured tobacco leaves. Zhang et al.(1997) proposed a method to grade tobacco leaves where mean of hue and chroma and standard deviation of hue are extracted as color features, and leaf length, leaf width, relation between leaf length and leaf width, average width of vein, ratio of length of vein and length of leaf are extracted as shape features. The features are represented in 2D-feature space and nearest neighbor rule are used to grade the tobacco leaves. Zhang et al., have used dataset of 110 cured tobacco leaves and achieved a classification accuracy of about 64%. A fuzzy classification system has been designed to grade tobacco leaves (Zhang et al., 2003) in which color, texture and shape features are used to design the model. Han (2008) in his work designed a model to grade the tobacco leaves where shape and color features are extracted and Support Vector Machine is used as a classifier to grade the cured tobacco leaves according to the part of growth of flue-cured tobacco leaves based on fuzzy statistics model. Barrel theory based decision-making algorithm was developed to grade cured tobacco leaves using color intensity, length and waste tolerance. Huabo et al. (2009) have worked on a dataset of 50 samples of four grades of LL group (leaf lemon group) and achieved a recognition rate of 85% (Huabo et al., 2009). A transformation technique from RGB signals to the Munsell system for color analysis of tobacco leaves was developed. Zhang et al. (1998) have extracted Munsell color features viz., average hue and standard deviation of hue for grades in each small groups of cured tobacco leaves.

The recent developments in the area of symbolic data analysis have proven that the real life objects can be better described by the use of symbolic data, which are extensions of classical crisp data (Billard and Diday, 2006). Suppose we want to describe the fruit (e.g., apricots) produced by a certain orchard. For example the weight is between 30 and 50 grams and its color is orange or green, but when the color is green then the weight is less than 35 grams (i.e., it's not yet ripe for eating). It is not possible to put this kind of information in a standard classical data table, since such tables consist of cells with a single point value. Nor is it easy to represent rules in a classical data table. This apricot is more easily represented by the description

[Weight = {30, 50}], [Color = {orange, green}] and [If {Color=green} then {Weight < 35].

This is a description associated with the concept of an apricot. The random variable values (Weight, Color) = ({30, 50}, {orange, green}) is an example of a two-dimensional symbolic-valued random variable. In general, a symbolic data 'point' assumes values as a hypercube (or hyper rectangle, broadly defined) or a Cartesian product of distributions in p-dimensional space or a mixture of both, in contrast to a classical data 'point' taking a single point value in p-dimensional space. Hence symbolic data may appear in the form of continuous ratio, discrete absolute interval and multi-valued, multi-valued with weightage, quantitative, categorical, etc. They result after aggregating a base dataset over individual entries that together constitute a second higher level (or category) of interest to the researcher (Billard and Diday, 2006).

The key idea of symbolic data analysis (SDA) originated in four century B.C. The Aristotle Organon (1994) clearly distinguishes first order individuals (as a horse or a man) considered as a unit associated to an individual of the world, from second order individuals (as the horse or the man) also taken as a unit associated to a class of individuals. For instance, in a census of a country, each individual of each region is described by a set of numerical or categorical variables given in several relations of a database. Such an individual is considered as first order individual. In order to study the regions considered as second order individuals, each variable can be considered as a random variable. Hence, each region in summarizing the values taken by its inhabitants can be described by inter-quartile intervals, or subsets of categorical values, or histograms or probability distributions, etc., depending on the concerned random variable. In such a way, symbolic data table is obtained where each row defines the description of a region and each column is associated with a symbolic variable. An extension of the standard data analysis to this type of data table is called symbolic data analysis (Diday (2002), Billard and Diday (2000)).

In cured tobacco leaves representation, since sample cured tobacco leaves of each grade posses significant variations, features extracted from such samples too vary considerably. Therefore, it would be more meaningful to capture these variations in the form of interval-valued features and to provide an effective representation for cured tobacco leaves. To the best of our knowledge, no work has been reported in literature, which uses symbolic representation for grading of cured tobacco leaves. Guru and Prakash (2009) made an attempt towards application of symbolic representation for signatures. While fixing inter valued feature they considered average and standard deviation of feature values whereas in this work, minimum and maximum of feature values are considered.

The organization of the paper is as follows, section 2 presents the proposed model which includes feature extraction and symbolic representation. In section 3 dataset and experimental results obtained using the proposed model are presented. The paper is concluded in section 4.

2 Proposed model

The proposed model has three stages, feature extraction followed by symbolic representation of cured tobacco leaves and finally symbolic matching for grading. Color features are extracted from cured tobacco leaves using Munsell system.

2.1 Feature extraction

Feature extraction is the process of transforming large input data into subset of features. If the features extracted are carefully chosen it is expected that the features set will extract the relevant information from the input data in order to perform the desired task using this reduced representation instead of the full size input. In the following paragraph, the input image data of a cured tobacco leaves are transformed into two color features to represent grades of cured tobacco leaves.

Density of blemish is very low in higher grades (grade 1 and grade 2) of cured tobacco leaves whereas it is high in lower grades (grade 3 and grade 4) and is reflected in the hue of leaves as shown in figure 1. The hue distribution of tobacco leaves is varying from one grade to another grade of cured tobacco leaves. Therefore proposed work exploited this by extracting color features - mean and standard deviation of hue of cured tobacco leaves using Munsell color system (Viscarra et al., 2006). Munsell color system encompasses the possible range of colors of cured tobacco leaves in different grades (Grade 1 to Grade 4) in each XL, LL and TL subgroups. Hue, value and chroma are the three variables of Munsell sytem. These variables describe a perceptual color space and not a quantitative measure of visible light. Hue is denoted by the letter abbreviation of the color of the spectrum (R for red, YR for yellow-red, Y for yellow) preceded by numbers from 0 to 10. Within each letter range, the hue becomes more yellow and less red as the numbers increase. The system was designed to arrange colors according to equal intervals of visual perception, thus the primary advantage of the Munsell system is its ease of interpretation. Munsell soil color book (Munsell, 1915), corresponding to value, chroma combinations for hues between 5R and 5Y inclusive, were used as proxy for soil color and as the source for the transformations between color space models. Since hue for cured tobacco leaves is also between 5R and 5Y (Zhang et al., 1997) proposed work is used the same transformations between color models as used in soil color analysis (Viscarra et al., 2006). Since there is no direct relationship between RGB to Munsell color system, RGB signals are transformed into CIE XYZ signals. In 1931 the Commission Internationale de l'Eclairage (CIE) standardised colour order systems by specifyingthe light source, the observer and the methodology used to derive the values for describing colour. In this color system, Y represents the brightness (or luminance) of the colour, while X and Z are virtual (or not physically realisable) components of the primary spectra (Viscarra et al., 2006). A non-linear transform from CIE XYZ to Munsell HVC is performed as explained in Viscarra et al. (2006). The examples of feature vectors (of type crisp) of samples of grade X1L are tabulated in Table 1.

(a) (b)

(c) (d)

Figure 1: Images of grades in XL (lugs and cutters lemon) group: (a) X1L (b) X2L (c) X3L (d) X4L

Table 1: Examples of feature vectors of samples of grade 1 of XL (i.e., X1L) of cured tobacco leaves

Feature vectors of samples of grade 1 of XL (i.e., X1L) of cured tobacco leaves		
Samples	Mean of munsell hue of cured tobacco leaf (Feature 1)	Standard deviation of munsell hue of cured tobacco leaf (Feature 2)
T_1	23.9181	1.0398
T_2	23.7051	1.1527
T_3	23.6193	1.2077
T_4	23.6735	1.1597
T_5	23.7047	1.1892
T_6	23.1210	1.3949
T_7	23.3754	1.2273
T_8	24.1211	0.9453
T_9	23.7792	1.2200
T_{10}	23.3986	1.1981

2.2 Symbolic Representation

Cured tobacco leaf samples of each grade form a class. Features (Mean and standard deviation of hue) of sample cured tobacco leaves of a grade extracted using Munsell system as given in section 2.1 have considerable intra class variation. Using conventional data representation preserving these variations is difficult. Hence, in this paper proposed work is intend to use unconventional data processing called symbolic data analysis which has the ability to preserve the variations among the data more effectively. In this paper, symbolic representation (Guru and Prakash, 2009) has been exploited to capture these variations through their assimilation by the use of interval valued feature vector as follows. Min-Max representation is another way to fix an interval valued feature vector and presented below.

Let $[T_{1,} T_{2,} T_{3,} \ldots\ldots\ldots, T_n]$ be a set of n samples of a cured tobacco leaves of grade say G_j ; $j = 1, 2, 3, \ldots\ldots\ldots\ldots, N$ (N denotes number of grades) and $F_i = [f_{i1,} \; f_{i2,} \; f_{i3,\ldots\ldots\ldots\ldots} f_{im}]$ be the set of m features characterizing the cured tobacco leaf sample T_i of the grade G_j .

Each k^{th} feature value of the G_j grade is represented by the use of interval valued feature $[f_{j,k}^{-}, f_{j,k}^{+}]$, where

$$f_{j,k}^{-} = Min_{j,k} \;\; and \;\; f_{j,k}^{+} = Max_{j,k} \tag{1}$$

where

Min_{jk} is the minimum of the k^{th} feature values obtained from all n samples of the grade G_j .

$$\text{i.e.,} \quad Min_{j,k} = \min(f_{1k,} \; f_{2k,} \; \ldots\ldots\ldots, f_{nk}) \tag{2}$$

and

Max_{jk} is the maximum of the k^{th} feature values obtained from all n samples of the grade G_j .

$$\text{i.e.,} \quad Max_{j,k} = \max(f_{1k,} \; f_{2k} \; \ldots\ldots\ldots, f_{nk}) \tag{3}$$

Each interval $[f_{j,k}^{-}, f_{j,k}^{+}]$ representation depends on the minimum and maximum of respective individual features. The interval $[f_{j,k}^{-}, f_{j,k}^{+}]$ represents the upper and lower limits of a k^{th} feature value of a cured tobacco leaf.

Now, the reference cured tobacco leaf for the grade G_j is formed by representing each feature $(F_i = [f_{i1,} \; f_{i2,} \; f_{i3,\ldots\ldots\ldots\ldots} f_{im}])$ in the form of an interval and is given by

$$RF_j = \{[f_{j1}^{-} \;\; f_{j1}^{+}][f_{j2}^{-} \;\; f_{j2}^{+}] \ldots\ldots\ldots [f_{jm}^{-} \;\; f_{jm}^{+}]\} \tag{4}$$

This symbolic feature vector is stored in the knowledgebase as a representative of the j^{th} grade. Similarly, symbolic feature vectors are computed for all individual grades $(j = 1, 2, 3, \ldots\ldots\ldots, N)$ and store them in the knowledgebase for future grading purpose. Thus, the knowledgebase has N number of symbolic vectors each corresponding to a grade. The examples of reference feature vectors for X1L, X2L and X3L are tabulated in Table 2.

Table 2: Examples of reference feature vector

Grades	Feature 1	Feature 2
X1L	[23.1210 24.1210]	[0.9453 1.3949]
X2L	[22.0769 22.9861]	[1.3749 2.2283]
X3L	[21.1160 22.1782]	[1.6186 2.6562]

2.3 Grading

In this section for grading the cured tobacco leaves, the work done by Guru and Prakash (2009) and Guru et al., (2010) is exploited. Symbolic classifier used for comparison purpose is as follows.

In our proposed grading model, the test sample of cured tobacco leaf is described by a set of m feature values of crisp type and compares it with the reference samples of all grades in the knowledgebase. Let $F_t = [f_{t1}, f_{t2}, f_{t3}, \ldots\ldots\ldots, f_{tm}]$ be a m dimensional feature vector representing a cured tobacco leaf. Let RF_j be a reference cured tobacco leaf for the j^{th} grade represented by an inter-valued feature vector as described in section 2.2. Cured tobacco grading strategy is to compare the test cured tobacco leaf F_t with all the reference cured tobacco leaves $RF_j,\ j = 1, 2, \ldots\ldots, N$ in the knowledgebase to obtain the A_C acceptance count for each reference cured tobacco leaf. The test cured tobacco leaf is said to belong to the grade with which it has a maximum acceptance count. Acceptance count A_C is given as below.

$$A_C = \sum_{k=1}^{m} C\,(f_{tk},[f_{tk}^{-}, f_{jk}^{+}])$$

where,

$$C\,(f_{tk},[f_{tk}^{-}, f_{jk}^{+}]) = \begin{cases} 1 & if\,(f_{tk} \geq f_{jk}^{-}\ and\ f_{tk} \leq f_{jk}^{+}) \\ 0 & otherwise \end{cases} \tag{5}$$

When the database happens to be large, there is possibility for a test cured tobacco leaf to possess the same maximum acceptance count with two or more reference cured tobacco leaves. Under such circumstances we recommend to resolve the conflict by the use of the following similarity measure (Guru and Prakash, 2009) which computes the similarity value between a test cured tobacco leaf and each of the conflicting grades say j^{th} grade.

$$Total\_Sim(F_t, RF_j) = \sum_{k=1}^{m} Sim(f_{tk},[f_{jk}^{-}, f_{jk}^{+}]) \tag{6}$$

Here $[f_{jk}^{-}, f_{jk}^{+}]$ represents the k^{th} feature interval of the j^{th} conflicting grade, and

$$Sim(f_{tk},[f_{jk}^{-},f_{jk}^{+}])=\begin{cases}1 & if\,(f_{tk}\geq f_{jk}^{-}\ and\ f_{tk}\leq f_{jk}^{+})\\ \max\left(\frac{1}{1+|22.9797-22.0761|*\delta},\frac{1}{1+|22.9797-22.9861|*\delta}\right)+1 & otherwise\end{cases} \tag{7}$$

where δ is a normalizing factor.

The examples for similarity value between feature vectors and reference feature vectors are tabulated in Table 3. Similarity value between feature vector (Feature 1 = 23.5316, Feature 2 = 1.2380) of test sample and reference feature vectors (Table 2) of grades X1L, X2L and X3L are 2, 0 and 0 respectively. Since both Feature 1 and Feature 2 values are lies in interval of respective reference feature vectors (Table 2) of X1L, similarity value (acceptance count) is 2 (equation (5)). Since both Feature 1 and Feature 2 values are not lies in interval of respective reference feature vectors (Table 2) of X2L and X3L, similarity value (acceptance count) for both X2L and X3L is 0 (equation (5)). Therefore feature vector (Feature 1 = 23.5316, Feature 2 = 1.2380) of test sample is assigned to the grade XIL.

Similarly Similarity value between feature vector (Feature 1 = 22.9797, Feature 2 = 1.6830) of test sample and reference feature vectors (Table 2) of grades X1L, X2L and X3L are calculated using equation 5. The similarity value (acceptance count) of grade X1L is 0. But similarity value (acceptance count) for grades X2L and X3L is 1. Therefore the conflict between grades X2L and X3L is resolved using equation 6. The calculated similarity values for X2L and X3L using equation 6 are 1.9936 and 1.5550 respectively and tabulated in Table 3. Therefore feature vector (Feature 1 = 22.9797, Feature 2 = 1.6830) of test sample is assigned to the grade X2L.

Similarity value between test sample feature vector (Feature 1 = 22.9797, Feature 2 = 1.6830) and **the grade X2L** (X2L Reference vector RF = {[22.0769 22.9861], [1.3749 2.2283]}) is calculated as follows.

$$Total\_Sim = \max\left(\frac{1}{1+|22.9797-22.0761|*\delta},\frac{1}{1+|22.9797-22.9861|*\delta}\right)+1$$

(Using equation 6 and equation 7)

Here δ is a normalizing factor whose value is set to 1.

$$Total\_Sim = \max(0.52255, 0.9936)+1$$

$$= 1.9936$$

Similarity value between test sample feature vector (Feature 1 = 22.9797, Feature 2 = 1.6830) and **the grade X3L** (X3L Reference vector RF = {[21.1160 22.1782], [1.6186 2.6562]}) is calculated as follows.

$$Total\_Sim = \max\left(\frac{1}{1+|22.9797-21.1160|*\delta}, \frac{1}{1+|22.9797-22.1782|*\delta}\right) +1$$

(Using equation 6 and equation 7)

Here δ is a normalizing factor whose value is set to 1.

$$Total\_Sim = \max(0.3491, 0.5550) +1$$
$$= 1.5550$$

Table 3: Examples for similarity value

Feature vectors of testing samples		Similarity value between feature vectors of testing samples with reference feature vectors of		
Feature 1	Feature 2	X1L	X2L	X3L
23.5316	1.2380	02	00	00
23.3425	1.3630	02	00	00
22.9797	1.6830	0	1.9936	1.5550
22.6773	1.6688	00	02	01
22.3308	1.8390	00	02	01
22.3604	1.7051	00	02	01
22.0074	2.0994	00	01	02
21.8121	2.1741	00	01	02
21.4327	2.3892	00	00	02

3 Experimental Results

3.1 Dataset

Color images of cured tobacco leaves are acquired using a Sony digital color camera in grading room at Central Tobacco Research Institute (CTRI), Hunsur, Karnataka, India. Images are acquired in an uncontrolled illumination. The data set contains X, L and T groups of cured tobacco leaves. Among these three groups our dataset includes only lemon color category forming XL, LL and TL subgroups of all four grades. The dataset constitute 887 samples of image size 350×150. Table 4 gives the number of samples of individual grades of each subgroup.

Table 4: Number of samples of individual grades of each subgroup

Grades	Subgroups			Over all total
	XL	LL	TL	
1	80	83	102	
2	75	127	130	
3	50	83	37	
4	40	40	40	
Total	245	309	333	887

3.2 Results

In order to corroborate the efficacy of the proposed method the effect of grading accuracy under varying size of database has been studied. Training set has been varied by 20%, 30%, 40%, 50% and 60% and remaining is used as testing. Samples are selected randomly for the purpose of training and testing. The grading accuracies under varying size of database of sub groups XL, LL and TL for four different grades are tabulated in Table 4, Table 5 and Table 6 respectively. Table 4, Table 5 and Table 6 gives the percentage of accuracy based on the types of grades. To study the overall performance of the system average grading accuracy of all grades of each sub groups has been calculated. From Table 7 it is clear that for 40% of training and 50% of training, best accuracy has been achieved on our real dataset.

Table 4: Grading accuracy for cured tobacco leaves of XL subgroup

Training samples	X1L	X2L	X3L	X4L	Average accuracy
20%	98.4375	80	48.3571	96.1538	80.7446
30%	98.0952	78.0952	50.6912	98.9011	81.4755
40%	100	73.3333	86.0215	96.1538	88.8772
50%	100	82.6667	83.8710	100	91.6344
60%	100	83.3333	80.6452	96.1538	90.0331

Table 5: Grading accuracy for cured tobacco leaves of TL subgroup

Training samples	T1L	T2L	T3L	T4L	Average accuracy
20%	99.3976	85.6299	36.1446	97.9730	79.7863
30%	99.8279	83.2396	92.9432	96.5251	93.1139
40%	98.3936	81.3648	90.0991	99.0991	92.3047
50%	98.7952	77.5663	91.5663	97.2973	91.2060
60%	99.3976	82.6772	90.3614	94.5946	91.7577

Table 6: Grading accuracy for cured tobacco leaves of LL sub group

Training samples	L1L	L2L	L3L	L4L	Average accuracy
20%	92.5926	65.6168	93.0931	99.0991	87.6004
30%	91.0364	77.6153	88.8031	96.5251	88.4950
40%	98.0392	77.4278	94.5946	99.0991	92.2902
50%	98.0392	75.5906	91.8919	97.2973	90.7047
60%	95.8820	82.6772	87.8378	94.5946	90.1745

Table 7: Grading Accuracy

Percentage of training	Average Accuracy of XL, LL and TL
20	82.7104
30	87.6948
40	91.1574
50	91.1817
60	90.6551

Also proposed model is compared with other state of the art models to study the elegance of the proposed model. The qualitative comparative analysis of the proposed model with other models is given in Table 8. From Table 8 it is understand that proposed model performs better than Zhang et al.(1997) in grading accuracy where as when compared with (Han, 2008) the grading accuracy is slightly less however the proposed model is competitive with it.

Table 8: Qualitative Comparison with other well known models on cured tobacco grading

Title	Grades	Size	Features	Classifiers	Accuracy in %
A trainable grading system for tobacco leaves (Zhang et al., 1997)	4 grades in each sub groups XL (lugs lemon), XF (lugs orange). 3 grades in each sub groups CL (cutters lemon), CF (cutters orange). 4 grades in each sub groups LL (leaf lemon), LF (leaf orange). 3 Grades in LR (leaf mahgony) Toatl Numer of grades = 23	110	Color features: Mean and standard deviation of hue, Mean chroma Texture feature : Roughness using sobel operator Shape feature: Ratio leaf length and leaf width, Vein width and length	Nearest Neighbor classifier	64%
Recognition of the part of growth of flue-cured tobacco leaves on suport vector machine (Han, 2008)	3 large groups X (lugs), C(cuters) and L(leaf) Total number of grades = 3	1712	Shape features: Width length ratio, Color features : hue, saturation and lightness	Support Vector Machine	94.62

Proposed model	4 grades in each sub groups XL (lugs and cutters lemon), LL (leaf lemon) and TL (tips lemon) Toatl Numer of grades = 12	887	Munsell Color features : Mean hue and Standard deviation of hue	Symbolic classification	91.1817

4 Conclusion

Proposed model has superior performance compared to other existing models for grading cured tobacco leaves. Proposed model has achieved good average classification accuracy of about 91%. Unconventional data representation called interval valued symbolic representation which helped in achieving good average classification accuracy for grading tobacco leaves. In future we will extend this work for all other standard grades and larger data set.

References

Aristotle (IV BC). (1994). Organon. Vol. I categories, II de 1'interpretation, J. Vrin edit. Paris.

Billard, L.and Diday, E. (2000). Regression analysis for inter-valued data. Proceedings of the seventh International Federation of Classification Societies: Data analysis, classificationand related methods (IFCS 200), July 11-14, Springer-Verlag.

Billard, L.and Diday, E. (2006). *Symbolic data analysis: Conceptual statistics and data mining.* Wiley series in computational statistics.

Diday (2002). An introduction to symbolic data analysis and sodas software. *Electro. J. Symb. Data Anal.* **0** (**0**), 1-25.

Guru, D.S. and Prakash, H.N. (2009). On-line signature verification and recognition: An approach based on symbolic representation. *IEEE Transaction on Pattern Analysis and Machine Intelligence* **31**, 1059- 1073.

Han, L. (2008). Recognition of the part of growth of flue-cured tobacco leaves based on support vector machine. *World Congress on Intelligent Control and Automation.* pp. 3624-3627.

Huabo, L, Liyuan, H. and Tao, L. (2009). The barrel theory based decision-making algorithm and its application. *International Conference on Computional Intelligence and Natural computing*. **1**, 11-14.

Manickavasagam, A, John, J., Gunasekaran and Doraisamy, P. (2007). Trends in Indian flue cured virgina tobacco (*Nictoina tobaccum*) processing: Harvesting, Curing and Grading. *Res. J. Agri. Biol. Sci*. **3**(**6**), 676-681.

Munsell, A. H. (1915). Munsell soil color charts. Munsell Color Company, Revised Edition, Macberth Division of Kollmorgen, New Windsor: NY.

Viscarra, R.A., Minasny, B., Roudier, P. and McBratney, A.B. (2006). Colour space models for soil science. *Geoderma* **133**, 320-337.

Zhang, H., Han, L. and Wang, Z. (2003). A fuzzy classification system and its application. *International Conference on Machine learning and cybernetics*. pp. 2-5.

Zhang, J., Sokhansanj, S., Wu, S., Fang, R., Yang, W. and Winter, P. (1997). A trainable grading system for tobacco leaves. *Comp. Electro. Agri*. **16**(3), 231-244.

Zhang, J., Sokhansanj, S., Wu, S., Fang, R., Yang, W. and Winter, P. (1998). A transformation technique from RGB signals to the Munsell system for color analysis of tobacco leaves. *Comp. Electro. Agri.* **19**, 155-166.

Author for correspondence

P B Mallikarjuna
Department of Studies in Computer Science, Manasagangothri,
University of Mysore,
Mysore-570006, Karnataka, India
E-mail: pbmalli@yahoo.com