Surface Grading Using Soft Colour-Texture Descriptors

Fernando López, José-Miguel Valiente, and José-Manuel Prats

Universidad Politécnica de Valencia, Camino de Vera s/n, 46022 Valencia, Spain
flopez@disca.upv.es

Abstract. This paper presents a new approach to the question of surface grading based on *soft colour-texture descriptors* and well known classifiers. These descriptors come from global image statistics computed in perceptually uniform colour spaces (CIE Lab or CIE Luv). The method has been extracted and validated using a statistical procedure based on *experimental design* and *logistic regression*. The method is not a new theoretical contribution, but we have found and demonstrate that a simple set of global statistics softly describing colour and texture properties, together with well-known classifiers, are powerful enough to meet stringent factory requirements for real-time and performance. These requirements are on-line inspection capability and 95% surface grading accuracy. The approach is also compared with two other methods in the surface grading literature; colour histograms [1] and centile-LBP [8]. This paper is an extension and in-depth development of ideas reported in a previous work [11].

1 Introduction

There are many industries manufacturing flat surface materials that need to split their production into homogeneous series grouped by the global appearance of the final product. These kinds of products are used as wall and floor coverings. Some of them are natural products such as marble, granite or wooden boards, and others are artificial, such as ceramic tiles. At present, the industries rely on human operators to carry out the task of surface grading. Human grading is subjective and often inconsistent between different graders [7]. Thus, automatic and reliable systems are needed. Capacity to inspect overall production at on-line rates is also an important consideration.

In recent years many approaches to surface grading have been developed (see Table 1). Boukouvalas et al [1][2][3] proposed colour histograms and dissimilarity measures of these distributions to grade ceramic tiles.

Other works consider specific types of ceramic tiles; *polished porcelain* tiles, which imitate granite. These works include texture features. Baldrich et al [4] proposed a perceptual approximation based on the use of discriminant features defined by human classifiers at factory. These features mainly concerned grain distribution and size. The method included grain segmentation and features

M. Lazo and A. Sanfeliu (Eds.): CIARP 2005, LNCS 3773, pp. 13–23, 2005.

measurement. Lumbreras et al [5] joined colour and texture through multiresolution decompositions on several colour spaces. They tested combinations of multiresolution decomposition schemes (Mallat's, *àtrous* and wavelet packets), decomposition levels and colour spaces (Grey, RGB, Otha and Karhunen-Loève transform). Peñaranda et al [6] used the first and second histogram moments of each RGB space channel.

Kauppinnen [7] developed a method for grading wood based on the Percentile (or centile) features of histograms calculated for RGB channels. Kyllönen et al´s [8] approach used colour and texture features. They chose centiles for colour, and LBP (Local Binary Pattern) histograms for texture description.

Lebrun and Macaire [9] described the surfaces of the Portuguese "Rosa Aurora" marble using the mean colour of the background and mean colour, absolute density and contrast of marble veins. They achieved good results but their approach is very dependent on the properties of this marble. Finally, Kukkonen et al [10] presented a system for grading ceramic tiles using spectral images. Spectral images have the drawback of producing great amounts of data.

Table 1. Summary of surface grading literature

	ground truth	**features**	**time study**	**accuracy %**
Boukouvalas	ceramic tiles	colour	no	-
Baldrich	polished tiles	colour/texture	no	92.0
Lumbreras	polished tiles	colour/texture	no	93.3
Peñaranda	polished tiles	colour/texture	yes	-
Kauppinen	wood	colour	yes	80.0
Kyllönen	wood	colour/texture	no	-
Lebrun	marble	colour/texture	no	98.0
Kukkonen	ceramic tiles	colour	no	80.0

Many of these approaches specialized in a specific type of surface, others were not accurate enough, and others did not take into account time restrictions in a real inspection at factory. Thus, we think surface grading is still an open research field and in this paper present a generic method suitable for use in a wide range of random surfaces. The approach uses what we call *soft colour-texture descriptors*, which are simple and fast [to compute] global colour and texture statistics. The method achieves good results with a representative data set of ceramic tiles. Furthermore, the approach is appropriate for use in systems with real-time requirements.

The final approach based on soft colour-texture descriptors (the proposed method) was extracted from a statistical procedure used to determine the best combination of quantitative/categorical factors in terms of a set of experiments that maximize or minimize one response variable also involved in the experiments. We used the accuracy rate of classifications as response variable. The statistical procedure is a combination of experimental design [13] and logistic regression [14] methods which have also been used for the literature approaches.

2 Literature Methods

For comparison purposes we selected two methods from the literature: colour histograms [1] and centile-LBP [8]. They are similar to ours, both are generic solutions with low computational costs. Colour histograms are 3D histograms (one axis per space channel) which are compared using dissimilarity measures. In [1] the authors used the *chi square test* and the *linear correlation coefficient.*

$$\chi^2 = \sum_i \frac{(R_i - S_i)^2}{R_i + S_i} \qquad\qquad r = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})}\sqrt{\sum_i (y_i - \bar{y})}}$$

When comparing two binned data sets with the same number of data points the *chi square* statistic (χ^2) is defined as above, where R_i is the number of events in bin i for the first data set, and S_i is the number of events in the same bin for the second data set. The *linear correlation coefficient* (r) measures the association between random variables for pairs of quantities (x_i,y_i), $i = 1,\ldots,N$. The mean of the x_i values is $\bar{x}$ and $\bar{y}$ is the mean of the y_i values.

The Centiles, are calculated from a cumulative histogram $C_k(x)$, which is defined as a sum of all the values smaller than x or equal to x in the normalized histogram $P_k(x)$, corresponding to the colour channel k. The percentile value gives x when $C_k(x)$ is known, so an inverse function of $C_k(x)$ is required. Let $F_k(y)$ be the percentile feature, then $F_k(y) = C_k^{-1}(y) = x$, where y is a value of the cumulative histogram in the range [0%,100%].

The Local Binary Pattern (LBP) is a texture operator where the original 3x3 neighbourhood is thresholded by the value of the centre pixel (Figure 1b). Pixel values in the thresholded neighbourhood are multiplied by the weights given to the corresponding pixels (Figure 1c). Finally, the values of the eight pixels are summed to obtain the number of this texture unit. Using LBP there are 2^8 possible combinations of texture numbers, then a histogram collects the LBP texture description of an image.

6	5	3
7	5	2
9	3	7

(a)

1	1	0
1		0
1	0	1

(b)

1	2	4
8		16
32	64	128

(c)

1	2	0
8		0
32	0	128

(d)

LBP=1+2+8+32+128=171

Fig. 1. Computation of local binary pattern (LBP)

In [8] centile and LBP features were combined in one measure of distance and then used the k-NN classifier. The Euclidean distance in the feature space was used for centile features. For LBP they used a log-likelihood measure: $L(S,R) = -\sum_{n=0}^{N-1} S_n ln R_n$, where N is the number of bins. S_n and R_n are the sample and reference probabilities of bin n. The distances were joined by simply adding them. Previously both distances were normalized using the min and max values of all the distances found in the training set.

3 Soft Colour-Texture Descriptors

The presented method is simple, a set of statistical features describing colour and texture properties are collected [15]. The features are computed in a perceptually uniform colour space (CIE Lab or CIE Luv). These statistics form a feature vector used in the classification stage where the well known k-NN and leaving-one-out methods [16] were chosen as classifiers.

CIE Lab and CIE Luv were designed to be perceptually uniform. The term 'perceptual' refers to the way that humans perceive colours, and 'uniform' implies that the perceptual difference between two coordinates (two colours) will be related to a measure of distance, which commonly is the Euclidean distance. Thus, colour differences can be measured in a way close to the human perception of colours. These spaces were chosen to provide accuracy and perceptual approach to colour difference computation. As the data set images were acquired originally in RGB, conversion to CIE Lab or CIE Luv coordinates was needed. This conversion is performed using the standard RGB to CIE Lab and RGB to CIE Luv transformations [17] as follows.

RGB to XYZ:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.412453 & 0.357580 & 0.180423 \\ 0.212671 & 0.715160 & 0.072169 \\ 0.019334 & 0.119193 & 0.950227 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

XYZ to CIE Lab:

$$L = 116(Y/Y_n)^{1/3} - 16$$
$$a = 500((X/X_n)^{1/3} - (Y/Y_n)^{1/3})$$
$$b = 200((Y/Y_n)^{1/3} - (Z/Z_n)^{1/3})$$

XYZ to CIE Luv:

$$L = 116(Y/Y_n)^{1/3} - 16$$
$$u = 13L(u' - u'_n)$$
$$v = 13L(v' - v'_n)$$

where

$$u' = 4X/X + 15Y + 3Z \qquad v' = 9X/X + 15Y + 3Z$$
$$u'_n = 4X_n/X_n + 15Y_n + 3Z_n \qquad v'_n = 9X_n/X_n + 15Y_n + 3Z_n$$

X_n, Y_n, and Z_n are the values of X, Y and Z for the illuminant (reference white point). We followed the ITU-R Recommendation BT.709, and used the illuminant D_{65}, where $[X_n\, Y_n\, Z_n] = [0.95045 \;\; 1 \;\; 1.088754]$.

We proposed several statistical features for describing surface appearance. For each channel we chose the mean and the standard deviation. Also, by computing the histogram of each channel, we were able to calculate histogram moments. The nth moment of z about the mean is defined as

$$\mu_n(z) = \sum_{i=1}^{L} (z_i - m)^n p(z_i)$$

where z is the random variable, $p(z_i)$, $i = 1, 2, \ldots, L$ the histogram, L the number of different variable values and m the mean value of z.

Colour histograms can easily collect 80,000 bins (different colours) which are all used to compute histogram dissimilarities. Centile-LBP approach uses 171 centile measures to compile colour property, and LBP histograms of 256 components to collect texture property. We can consider that these approaches use 'hard' colour and texture descriptors in comparison to our method which only uses the mean, standard deviation and histogram moments from 2nd to 5th to compile colour and texture properties (a maximum feature vector of 18 components). By comparison we named the proposed method *soft colour-texture descriptors*. This assertion is even more acceptable if we revise classical approaches to texture description in the literature.

4 Experiments and Results

All the experiments were carried out using the same data set. The ground truth was formed by the digital RGB images of 960 tiles acquired from 14 different models, each one with three different surface classes given by specialized graders at factory (see Table 2 and Figure 2). For each model two classes were close and one was far away. Models were chosen representing the extensive variety that factories can produce, a catalogue of 700 models is common. But, in spite of this great number of models, almost all of them imitate one of the following mineral textures; marble, granite or stone.

Table 2. Ground truth formed by 14 models of ceramic tiles

	classes	**tiles/class**	**size (cm)**	**aspect**
Agata	13, 37, 38	16	33x33	marble
Antique	4, 5, 8	14	23x33	stone
Berlin	2, 3, 11	24	16x16	granite
Campinya	8, 9, 25	30	20x20	stone
Firenze	9, 14, 16	20	20x25	stone
Lima	1, 4, 17	24	16x16	granite
Marfil	27, 32, 33	14	23x33	marble
Mediterranea	1, 2, 7	30	20x20	stone
Oslo	2, 3, 7	24	16x16	granite
Petra	7, 9, 10	28	16x16	stone
Santiago	22, 24, 25	28	19x19	stone
Somport	34, 35, 38	28	19x19	stone
Vega	30, 31, 37	20	20x25	marble
Venice	12, 17, 18	20	20x25	marble

Digital images of tiles were acquired using a spatially and temporally uniform illumination system. Spatial and temporal uniformity is important in surface grading [1,4,6] because variations in illumination can produce different shades for the same surface and then misclassifications. The illumination system was formed by two special high frequency fluorescent lamps with uniform illuminance

Fig. 2. Samples from the ground truth. From up to down; three samples of petra and marfil models, each one corresponding to a different surface grade.

along their length. To overcome variations along time, the power supply was automatically regulated by a photoresistor located near the fluorescents.

In order to study the feasibility of the soft colour-texture descriptors on perceptually uniform colour spaces we carried out a statistical experiment design. Our aim was to test several factors to determine which combination gave the best accuracy results. These factors related to colour spaces, classifiers, and sets of soft colour-texture descriptors. Colour space: CIE Lab, CIE Luv, RGB and Grey scale. Classifier: k-NN (k=1,3,5,7) and leaving-one-out (k=1,3,5,7). Soft colour-texture descriptors: mean, standard deviation and 2nd to 5th histogram moments.

The factors and their possible values defined 4096 different classification experiments for each tile model. As the ground truth was formed by 14 tile models, 57,344 experiments had to be carried out. We decided to use a statistical tool, the *experiment design* [13], in order to manage the large quantity of experiments and results. This tool, in combination with the *logistic-regression* [14] method, provides a methodology for finding the best combination of factors in a set of experiments that maximize or minimize one response variable. In our case, we were looking to maximize classification accuracy rates. This methodology follows the plan presented in Figure 3.

When we want to perform a complex experiment or set of experiments efficiently we need a scientific approach to planning the experiment. Experimental design is a statistical tool which refers to the process of planning experiments so that appropriate data can be collected for analysis with statistical methods. This would lead to objective and valid conclusions. We chose to use a complete factorial approach in the design of our experiment. This is the most advisable approach for dealing with several factors [13]. Complete factorial design means

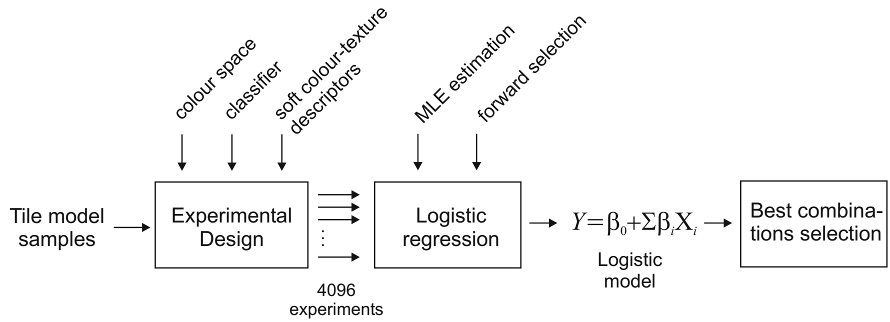

Fig. 3. Block diagram representing the statistical procedure for extracting the best combinations of factor values in a set of experiments or experimental design

that we select a fixed number of possible values for each factor and then carry out experiments with all the possible combinations of them. In our case, each combination of factors is a single experiment with a classification of surface grades. By varying the factor values in a nested way independence between factors, iterations and experiments is achieved, guaranteeing that simple and interaction effects are orthogonal.

From the set of performed experiments we computed the logistic model using a logistic regression [14]. The achieved mean accuracy of all models is used as the output variable (response variable). Thus, we summarize the 14 groups of 4096 experiments for tile models in a single set of 4096 experiments. We used a logistic (logarithmic) approach rather than a linear one because the output variable is probabilistic and the logarithmic method fits the extracted model better. Using the extracted logistic model, $y = \beta_0 + \sum \beta_i X_i$, we compute the predicted accuracy rate for each combination of factors using $p = \frac{e^y}{1+e^y}$. Then we can sort the combinations by their predicted accuracies. The one with the best accuracy will reveal the best combination of factors.

The best predicted accuracy rate in the experimental design carried out for soft colour-texture descriptors was 97.36% with a confidence interval at 95% [96.25%, 98.36%]. This result was achieved using CIE Lab colour space, 1 leaving-one-out classifier and all the proposed soft colour-texture descriptors (mean, standard deviation and 2nd to 5th histogram moments). The measured accuracy with this combination was 96.7%.

Figure 4 shows that CIE Lab and CIE Luv spaces featured strongly in the best sets of factor combinations. RGB space achieved almost null presence in the 1000 best combinations rising to 11.9% and 16.3% in 1500 and 2000 combinations. The Grey scale (with no colour information) was not among the best combinations. Thus, perceptually uniform colour spaces show clearly better performance than RGB with the soft colour-texture descriptors method. Also, this figure suggests that best classifiers are derived from the leaving-one-out method.

A similar experimental design was performed for the literature methods. In this case, we used the following factors and possible values. Colour space: CIE

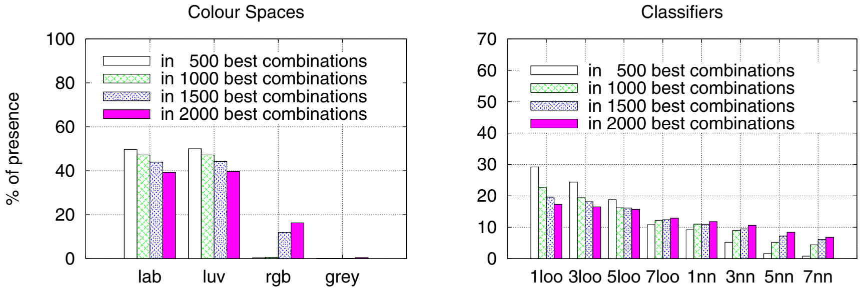

Fig. 4. Presence percentage of colour spaces and classifiers in the best combinations sets ordered by the predicted accuracy rate in the experimental design performed for soft colour-texture descriptors

Lab, CIE Luv, RGB and Grey scale. Classifier: k-NN (k=1,3,5,7) and leaving-one-out (k=1,3,5,7). Distance measure: chi square test, linear correlation coefficient, log-likelihood measure. Distance measures are used in these methods to determine colour histograms and LBP histograms dissimilarities. In a study similar to the one in Figure 4 it was concluded that RGB was the best space for the colour histograms approach closely followed by CIE Lab. Nevertheless, in centile-LBP, CIE Lab was the best followed by RGB. Chi square test was the best distance in colour histograms, and linear correlation performed better in centile-LBP. In both methods the leaving-one-out classifiers again showed the best performance. Table 3 shows the best results achieved in each surface grading method and its corresponding combination of factors.

In all methods the achieved performance is very good and quite similar. For all of them predicted accuracy and confidence interval exceed factory demands of 95%. Differences between the methods arose in terms of timing costs.

Figure 5 presents a comparison of the methods by timing costs (measured on a common PC) for nine of the fourteen tile models. The soft colour-texture

Table 3. Best result of each surface grading approach

	factors	**predicted accuracy**	**c.i. 95%**	**measured accuracy**
Soft colour-texture descriptors	CIE Lab, 1-loo, all descriptors	97.36%	[96.25%, 98.36%]	96.7%
Colour histograms	RGB, 1-loo chi square	97.82%	[96.50%, 98.54%]	98.67%
Centile-LBP	CIE Lab, 1-loo, linear correlation	98.26%	[97.27%, 99.03%]	98.25%

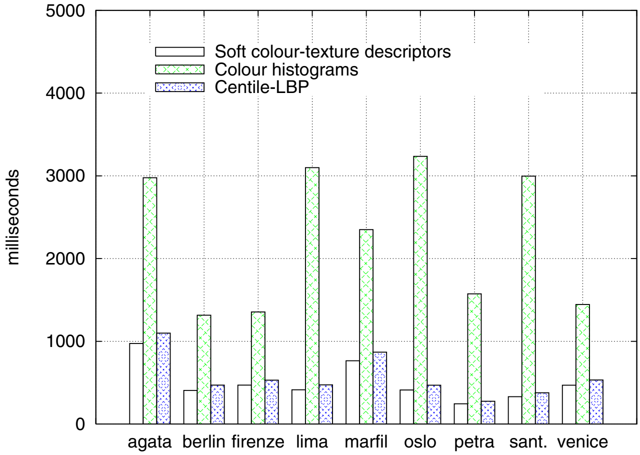

Fig. 5. Timing comparison of surface grading approaches using the corresponding best combination of factors in each method

descriptors method provides the best performance, closely followed by centile-LBP. The colour histograms approach compile by far the worst timing despite the fact that this method does not need to translate the image data, originally in RGB, into CIE Lab or CIE Luv spaces. Also, this method presents irregular timing for the same data size. The berlin, lima and oslo models share data size (tile and image size) but the method achieves significant timing differences among them. This effect is due to the use of binary trees to store the colour histograms of images. Images with larger numbers of different colours need larger trees and more time to compute the differences between histograms. This timing dependence related to data values does not appear in the other two methods whose computational costs only depend on image size and the algorithm; $\Theta(n) + C$ where n is the image size and C is a constant related to the algorithm used for implementing the approach.

5 Conclusions

In this paper we present a new approach for the purpose of surface grading. This approach is based on the use of soft colour-texture descriptors and perceptually uniform colour spaces. Two statistics tools, experimental design and logistic regression, has been used to study and determine the best combination of factors providing the best accuracy rates using a ground truth composed of 14 ceramic tile models. The best combination was: CIE Lab colour space, 1 leaving-one-out classifier and all the soft colour-texture descriptors.

For comparison purposes, a similar study was performed for two literature methods; colour histograms and centile-LBP. In this study we used the factor of inter-histograms distance measures instead of soft colour-texture descriptors.

Best combinations of factors were RGB colour space, 1 leaving-one-out classifier and chi square distance for the colour histograms method, and CIE Lab, 1 leaving-one-out classifier and linear correlation for centile-LBP.

All the approaches achieved factory compliance exceeding the 95% of minimum accuracy. The achieved percentages of all methods vary in less than 1%, thus the accuracy results are quite similar. The differences among the methods arose more clearly when we studied the timing costs. The best in timing was the method based on soft colour-texture descriptors closely followed by centile-LBP. Colour histograms performed worse and irregularly due to binary trees which are used to efficiently store the histograms.

In a work recently reported [12] we studied and demonstrate the on-line inspection capability of soft colour-texture descriptors carrying out a study of real-time comliance and parallelization based on MPI-cluster technology.

Acknowledgements

This work was partially founded by FEDER-CICYT (DPI2003-09173-C02-01). We also want to thank the collaboration of Keraben S.A. for providing the ceramic tile samples.

References

[1] C. Boukouvalas, J. Kittler, R. Marik and M. Petrou. Color grading of randomly textured ceramic tiles using color histograms. IEEE Trans. Industrial Electronics, 46(1):219–226, 1999.

[2] C. Boukouvalas and M. Petrou. Perceptual correction for colour grading using sensor transformations and metameric Data. Machine Vision and Applications, 11:96-104, 1998.

[3] C. Boukouvalas and M. Petrou. Perceptual correction for colour grading of random textures. Machine Vision and Appl., 12:129-136, 2000.

[4] R. Baldrich, M. Vanrell and J.J. Villanueva. Texture-color features for tile classification. EUROPTO/SPIE Conf. on Color and Polarisation Techniques in Indust. Inspection, Germany, 1999.

[5] F. Lumbreras, J. Serrat, R. Baldrich, M. Vanrell and J.J. Villanueva. Color texture recognition through multiresolution features. Int. Conf. on Quality Control by Artificial Vision. 1:114–121, France, 2001.

[6] J. A. Peñaranda, L. Briones and J. Florez. Color machine vision system for process control in ceramics industry. SPIE. 3101:182–192, 1997.

[7] H. Kauppinen. Development of a color machine vision method for wood surface inspection. Phd Thesis, Oulu University, 1999.

[8] J. Kyllönen and M. Pietikäinen Visual inspection of parquet slabs by combining color and texture. Proc. IAPR Workshop on Machine Vision Appl., Japan, 2000.

[9] V. Lebrun and L. Macaire. Aspect inspection of marble tiles by color line-scan camera. Int. Conf. on Quality Control by Artificial Vision, France, 2001.

[10] S. Kukkonen, H. Kvinen and J. Parkkinen. Color Features for Quality Control in Ceramic Tile Industry. Optical Engineering. 40(2):170–177, 2001.

[11] F. López, J.M. Valiente, R. Baldrich and M. Vanrell. Fast surface grading using color statistics in the CIE Lab space. 2nd Iberian Conference on Pattern Recognition and Image Analysis (IbPRIA2005). Lecture Notes in Computer Science. 3523:666-673, 2005.

[12] F. López, J.M. Valiente and G. Andreu. A study of real-time compliance and parallelization for the purpose of surface grading. XVI Jornadas de Paralelismo, I Congreso Español de Informática, 2005.

[13] D.C. Montgomery. Design and analysis of experiments. (4th Edition), John Wiley and Sons, Inc. New York, 1997.

[14] R. Christensen. Log-linear models and logistic regression. (2nd Edition), Springer-Verlag, New York, 1997.

[15] R.C. Gonzalez and P. Wintz. Digital image processing. Addison-Wesley, 2nd Edition, 1987.

[16] R.O. Duda and P.E. Hart. Pattern classification and scene analysis. John Wiley and Sons, New York, 1973.

[17] G. Wyszecki and W.S. Stiles. Color sciencie: concepts and methods, quantitative data and formulae. Wiley, 2nd Edition, New York, 1982.