Nonlinear Variance Measures in Image Data

Carolyn J. Evans and Imants D. Svalbe

Department of Physics, Monash University, Clayton 3168, Australia.
carolyn.evans@sci.monash.edu.au

Abstract. The homogeneity of regions in images can be measured in terms of the variation of local image values, or in terms of the local variation of ranks assigned to those image values. Previously proposed measures of nonlinear, or rank variance are shown in this paper to be insufficient measures of rank variation, especially when applied to discrete image data. A more useful measure of rank variance for image analysis, called diversity, is presented and characterised here. The dependence of diversity on the size of the data set involved and on the number of possible data values is discussed. The measure provides a very concise summary of the rank structure about each point and is sensitive to 'ties' in the local rank distribution.

Keywords : rank variance, non-parametric statistics, order statistics, robust estimation

1 Introduction

The homogeneity of an image region can be defined in different ways. For example, an image region may be considered relatively inhomogeneous when values vary greatly across the region and its linear variance is high. Alternatively, the homogeneity of an image region may be dependent on the number of tied values within the region. An inhomogeneous or 'rough' image region is one in which no two values are the same and no 'ties' exist. (Ties are not unlikely to occur in image data, which is often discrete.) Such a region should have high nonlinear, or rank variance, because then every value has a different rank relative to the local image data.

It is indeed suggested by Kim and Yaroslavskii [1], that the adaptive properties of algorithms based on ranks would be a convenient tool for measuring local statistical characteristics of pictures such as local mean, local variance and other distribution moments. However we can not measure image roughness defined in the nonlinear sense using previous definitions of rank variance because they are not based solely on rank, or they are insensitive to the presence of 'ties' in local image data. For example, R-estimates of variance, like the σ_R^2 mentioned by Draper [2], are not solely based on rank and have the same cardinal measurement scale as the original data.

Puri and Sen [3] define a nonlinear variance by defining covariance based on general rank scores. This nonlinear variance is insensitive to ties in the input

data distribution. It is assumed the input data distribution is continuous and ties are unlikely to occur. Draper also does not account for ties in the rank data because he considers only continuous data.

There is also a rank variance measure, associated with Spearman's rank correlation measure, ρ [4]. Spearman's ρ is designed to be insensitive to ties in the orderings of two ranked variables, by assigning midranks to tied values, since ties are not relevant to determining the rank correlation between the orderings. The associated rank variance is likewise insensitive to ties. The frequency distribution of of figure 1a, shows that most rank distributions have high normalised rank variance, although only one rank distribution does not contain any ties. This frequency distribution was produced by generating all of the possible 2^{N-1} rank distributions, for $N = 25$, with midranks assigned to ties. The number of ties increases along the vertical axis from 0 to 24 and the rank variance increases from left to right on the horizontal axis, with a range of 0 to 1.

In this paper we present and characterise an alternative rank variance, called diversity, which can be used to measure image roughness. In section 2 the diversity is shown to be sensitive to the presence and location of ties in local rank distribution of image data, because ranks are not assigned using the midrank method. In section 3 comparisons are made between using diversity and linear variance to measure image homogeneity and concluding comments are made in section 4.

2   Diversity, d

The diversity measures the rank variance of N data values, where N is odd. Ranks are assigned to the N data values using the same method as Chung and Marden [5]. These ranks are all integer values, and have a maximum range from 1 to N. The lowest data value in the sample is assigned the rank 1. The next lowest value is ranked 2 and so on. If ties exist, all tied values are ranked equally. For example the values, 0, 3, 3, 5, 10 would have the rank distribution 1, 2, 2, 3, 4.

Diversity is then defined as,

$$d = \frac{\sum_{i=1}^{N}(r_i - r_{med})^2}{N(N^2 - 1)/12},$$

(1)

where r_i is the assigned rank of the ith data value, r_{med} is the rank of the median data value, and $N(N^2 - 1)/12$ is a normalising factor.

The diversity of single-band image data is calculated using a moving, local window containing N pixels. The ranking process must be repeated every time new data enters the moving window before the diversity can be calculated.

2.1   Diversity versus Ties

The frequency distribution of diversity is shown in figure 1b for $N = 25$. This distribution was determined exactly by generating all the possible rank distributions for $N = 25$ and calculating the diversity for all possible rank distributions.

A comparison of the two curves of figures 1a and b indicates the diversity is more sensitive to the number of ties than Spearman's rank variance, as more mid-range values of diversity are possible for moderately 'rough' rank distributions. Diversity is also sensitive to the location of ties within the rank distribution.

2.2 Dependence of d on N

The diversity, d, is dependent on N, the size of the data sample. To show this, the probability distribution of d, for odd N between 7 and 121, were determined experimentally. In the cases where $N < 29$, all possible rank distributions were generated and the associated diversity was calculated. For $N \geq 29$, the number of possible rank distributions becomes very large. Instead, a sample of 10 million rank distributions were randomly generated. (It was found to be intractable to derive the theoretical probability distributions analytically, even for small N.)

For low N, the probability distribution of d does not vary smoothly, making it difficult to characterise. Instead the cumulative probability distributions of d were computed. Each cumulative probability distribution was characterised by fitting a curve of the form,

$$Pr(d \leq y) = A\{1 + (p-1)e^{-k(y-y_c)}\}^{1/(1-p)}, \tag{2}$$

where A, y_c, d and k were variable parameters. This curve was chosen because its shape was most similar to the form of the experimental data. It was fitted to minimise the mean square deviation of the curve from the data (using Microcal Origin). Figure 2a contains the experimental data and fitted curves for $N = 9$, 25, 49 and 121.

The variation of the parameters A, y_c, p and k with N are characterised by,

$$A = 1.001 + 0.023e^{(-(N-7)/5.3)} + 0.004e^{(-(N-7)/33)}, \tag{3}$$

$$y_c = 0.2486 - 0.0273e^{(-(N-7)/2.17)} - 0.0073e^{(-(N-7)/20)}, \tag{4}$$

$$p = 1.658 - 0.527e^{(-(N-7)/44)} - 0.356e^{(-(N-7)/5.4)}, \tag{5}$$

$$k = 39.7 - 34.3e^{(-(N-7)/95)}. \tag{6}$$

Curves (3)–(6) were fitted to minimise the sum of the squared deviations of the curve from the numeric data, weighted by the experimental error of each data point.

We characterised the dependence of the cumulative probability distributions of diversity on the sample size, N, so that comparisons could be made between measures of image roughness at different scales. Diversity measurements made at scale N_1 could be compared with diversity measurements made at scale N_2, where $N_1 > N_2$, through the cumulative probability of those diversity measurements at the respective scales.

2.3 Dependence of d on the Number of Possible Data Values

Up until now it has been assumed that image data can range over at least N different values, so that it is possible for no ties to occur in a sample of N

data points. However, the diversity of discrete image data does depend on the quantisation level of the data, by nature of our nonlinear definition of image roughness. If the data range contains only g different values, where $g < N$, a sample of N data points will contain at least $N - g$ ties. Figure 2b gives an example of the variation in the cumulative distribution of d for N=9 and $g = 9, 5$ and 2.

The cumulative distribution for $g = 2$, ie. binary image data, can be derived analytically. It is given by,

$$Pr(d \leq y) = \frac{1}{N} + \frac{2(N^2 - 1)y}{12}. \tag{7}$$

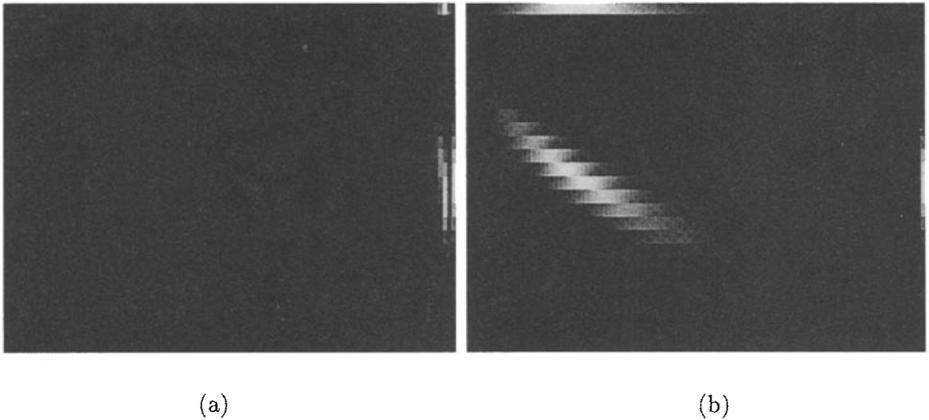<center>(a)   (b)</center>

Fig. 1. (a) Bivariate frequency histogram of the number of rank distributions ($N = 25$) with a certain number of ties and Spearman rank variance. (b) Bivariate frequency histogram of the number of rank distributions ($N = 25$) with a certain number of ties and diversity. (The number of ties increases along the vertical axis from 0 to 24, bottom to top, and the normalised rank variance or the diversity increase along the horizontal axis from 0 to 1, left to right.)

3 Comparing Linear Variance and Diversity

The insensitivity of the diversity measure to actual image data values can be advantageous. The linear variance and diversity have been calculated for the grey scale image of figure 3a using a 5×5 moving window. The results are given in figures 3b and 3c. Because linear variance is contrast dependent, regions of the image containing small fluctuations in data values, for example the hat band, are considered smooth compared to regions containing larger fluctuations
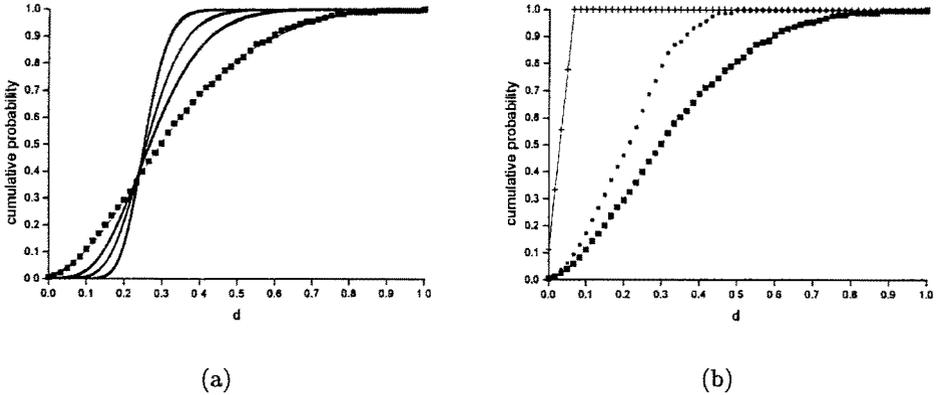
(a)   (b)

Fig. 2. (a) Cumulative Probability Distribution (CPD) of diversity for N=9, 25, 49, and 121, with fitted curves. (As N increases the curves become more steep), (b) CPD for d, at $N = 9$, and $g = 2$, 5, and 9, respectively

in data values. The most inhomogeneous regions are the edges of the hat and the shoulder. On the other hand, since diversity is contrast independent, the diversity of the hat band region is nearly as great as the edges of the hat and shoulder. However, the diversity of noise fluctuations in the image background is also more significant because of the associated fluctuations in local rank.

Linear variance does not have a fixed dynamic range because it is contrast dependent, so it is only a relative measure of roughness. The dynamic range of diversity is fixed by the size of window used to calculate diversity, and has been normalised. The fact that the quantisation of discrete image data affects the local diversity may not be a desirable in some applications. However if the image grey values have undergone some nonlinear mapping during data acquisition, linear variance is not a reliable measure of roughness, whereas diversity is invariant to any strictly order-preserving mapping applied to image values. This can be seen by comparing the results of figure 4 to those given in figure 3. Figure 4a is figure 3a after log contrast stretching. Figure 4b and 4c contain respectively the linear variance and the diversity of figure 4a, calculated using a 5 × 5 local window.

4 Conclusion

The measure of nonlinear variance, known as diversity, was presented in this paper. It is more sensitive to ties in image data than previous definitions of nonlinear, or rank variance. We have characterised the diversity so that it can be used as a measure of the inhomogeneity or 'roughness' of image regions over different scales.

Diversity could be used to highlight the presence of watermarks or signatures added to the low order bits of image pixels. Whilst the grey level changes of the added values will not be perceptible, the changes to the local ranks of pixels may

be significant, even if the signature bits are encrypted across edges or relatively noisy parts of the image.

The diversity has been applied to discrete image data in this paper, but we discuss how it should be applied to double-precision image data in [6]. The in-band diversity has also been extended to a cross-band 'co-diversity' for multi-band image data in [6].

Acknowledgements

References

1. Kim V. and Yaroslavskii L.. Rank Algorithms for Picture Processing. CVGIP **35** (1986), 234–258.
2. Draper D.. Rank-Based Robust Analysis of Linear Models.I. Exposition and Review. Statistical Science, **3**(2) (1988), 239–271.
3. Puri M.L.and Sen P.K.. Analysis of Covariance Based on General Rank Scores. Ann. Statistics, **40**(2) (1969), 610–618.
4. Kendall, M.G.. Rank Correlation Methods 4th Ed. Griffin, London, 1970.
5. Chung L. and Marden J.I.. Use of Nonnull Models for Rank Statistics in Bivariate, Two-Sample, and Analysis of Variance Problems. J. American Statistical Association, **86**(413) (1991), 188–200.
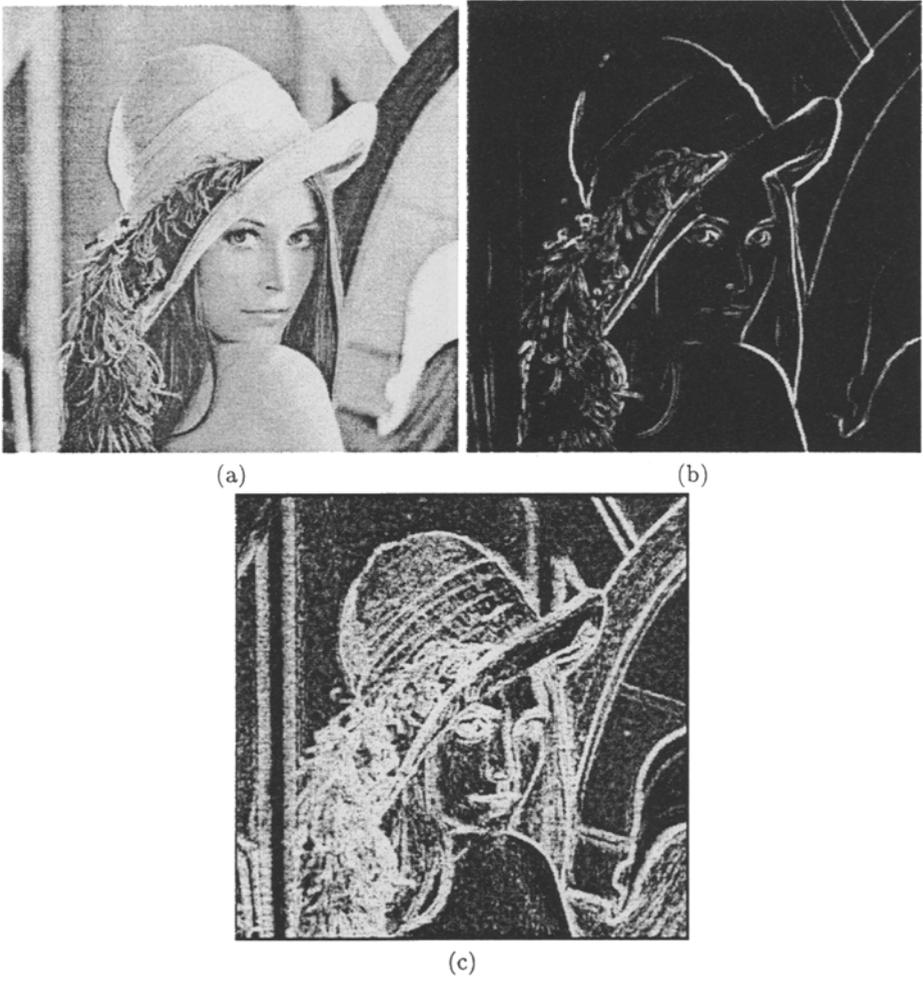6. Svalbe I.D. and Evans C.J.. Nonlinear Co-variance for Multi-band Image Data. To be presented at SSPR '98.

Fig. 3. (a) Original grey scale image of Lena. (b) Linear variance and (c) diversity calculated from (a) using a 5 × 5 window.
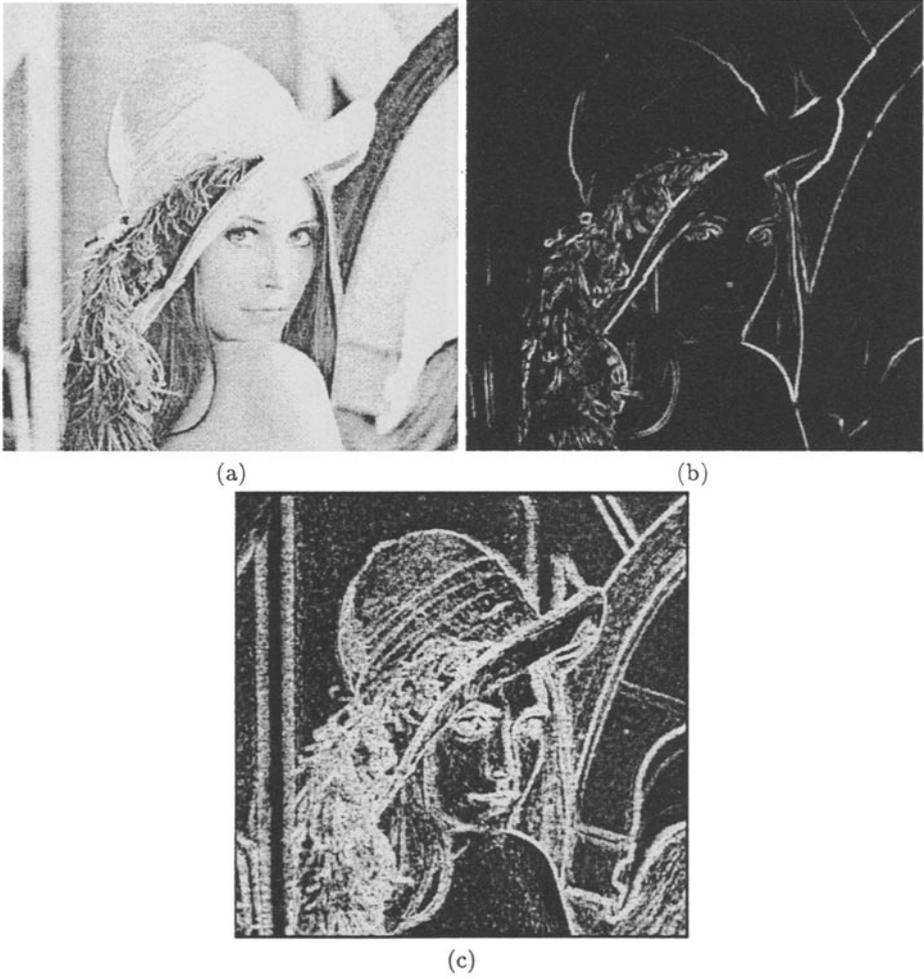
(a)

(b)

(c)

Fig. 4. (a) Log contrast stretched image of Lena. (b) Linear variance and (c) diversity calculated from (a) using a 5 × 5 window.