Tina Memo No. 2010-005
Internal, Literature review.

Automated Planetary Image Analysis and Associated Literature

Paul Tar

Last updated
20/04/2010

Imaging Science and Biomedical Engineering Division,
Medical School, University of Manchester,
Stopford Building, Oxford Road,
Manchester, M13 9PT.

Contents

1 Introduction

The planetary science community has been furnished with many thousands of surface images from our nearest planetary neighbours since the 1960s. Early datasets were both low in volume and in resolution making the job of their analysis tractable by human experts, despite it being a laborious and time consuming task. Recent missions have been equipped with high resolution imaging capabilities yielding many more images than human analysts could ever reasonably process manually. Massive Martian datasets generated from the Mars Reconnaissance Orbiter for example include images with resolutions up to 50m over large regions of the planet. Complementary to these images, 3D terrain maps exist for some areas notably from Mars Orbiter Laser Altimeter data, but these are of a lower resolution and are not available in all cases. The ever increasing number of available images and future missions such as DAWN to the asteroid belt and Messenger to Mercury add impetus to the search for a reliable automated assistant to aid planetary image analysis. As well as the potential labour savings, automation can help remove inherent subjectivity of human analysts thereby providing more consistent measurements to be extracted from images. This document reviews potential planetary science applications of an automated system, critically examines a number of existing algorithms for surface feature extraction and investigates possible scientifically valid solutions.

2 Applications

2.1 Crater counting

Chronological studies of planetary surfaces rely heavily on crater counting. The older a surface the more impact craters there are expected to be present. Debris in the early solar system is believed to have contained many large objects resulting in the biggest impact craters during the late heavy bombardment approximately 4.0GA ago [1]. After this period it is believed there has been a steady accumulation of smaller impacts up to the present day. A Size Frequency Distribution (SFD) is a plot relating the size and number of impact craters present within an area [2]. The shape of these plots is indicative of a surface's age. Radiometric dating techniques have been applied to lunar samples thereby providing absolute calibration for SFDs applicable to the Moon. Other calibrations not yet withstanding, SFDs can still be applied to give relative ages for Martian [3][4] and other surfaces. Figure 1 shows an example SFD. Beyond simply dating a surface, an SFD may be used to gain information about erosive and other processes acting upon a surface as craters are in-filled and weathered at different rates under different conditions.

2.2 Drainage networks

The study of drainage networks can provide information on climate history [5] and allow inferences to be made about tectonic modification of a planetary surface. The parameters of a drainage network can help determine the method by which the drainage system was created. Drainage density, defined as the total length of drainage channels divided by the area of a drainage basin, is expected to be high if networks are formed in the presence of precipitation and surface runoff. Conversely, lower drainage densities with channels exhibiting amphitheatre type terminations are better accounted for by groundwater sapping [12]. As water will always drain via the steepest path it is also possible to infer relative surface elevations from visual images. A great many drainage networks have been observed on Mars, predominantly, but not exclusively, in the cratered southern uplands. These are believed to have been formed in the Noachian period between 3.5GA and 4.6GA ago. As these networks have been inactive for billions of years any inferred elevations will reveal paleo-gradients which can be compared to present day gradients, discrepancies between which provide evidence of tectonic modification. As of yet no examples have been observed on Mars but Earth exhibits numerous uplifted networks [6] [7].

3 Published feature detection methods

3.1 Crater detection

A broad range of algorithms have been proposed for the extraction of impact crater structures in images. All algorithms face similar problems inherently found in crater images. Briefly, these problems include:

- Craters are different sizes

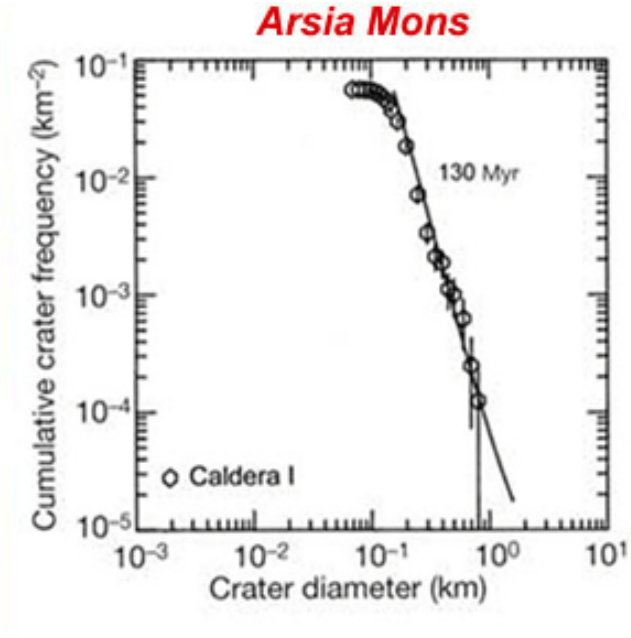

Figure 1: Size Frequency Distribution of impacts in the Arsia Mons caldera. Y axis shows cumulative frequency of craters, X axis shows diameter of craters. It is typical for both axis to be plotted on a logarithmic scale. The shape of the curve is indicative of the caldera's age.

- Camera angles can cause craters to appear elliptical
- Crater illumination can vary considerably
- Craters can overlap, partially destroying crater walls
- Craters can be nested
- There are many other structures which can appear crater like

No proposed algorithm fully addresses all of these problems, preferring to focus on simpler problems such as scale invariance. An overview of currently published crater detection methods is presented below.

3.1.1 Shadow topology method

Andre Smirnov of The University of Colorado developed a scale invariant crater detection algorithm based on shadow topology [9]. This method was implemented and tested using Sandia National Labs' FCDMF imaging library. The algorithm consists of 4 mandatory steps, followed by 2 optional stages purported to improve performance. Step 1 is the thresholding of an image in order to extract shadows. This step involves user intervention to subjectively select and experimentally adjust a threshold to extract as much genuine shadow as possible. Step 2 clusters connected putative shadow pixels into separate objects. Step 3 places a bounding rectangle around shadow objects and performs a simple filtering to eliminate objects deemed to be too large or small. Step 4, the last mandatory step, searches for sharp discontinuities around object perimeters to differentiate between crater and non-crater shadows. The discontinuity test operates under the premise that extremal points on crater shadows will form acute angles with their neighbours as illustrated in Figure 2. An object containing two such discontinuities is considered to be a crater.

The optional steps 5 & 6 are intended to remove false positives. A circle is fitted using three points on an object's perimeter, two of which being the discontinuity points and the third being selected at random. The mean distance of all perimeter pixels from the fitted circle is used to reject objects if they fall outside of a user defined tolerance. The author estimates that 80% positive detection rates are achievable and reports 100% successful detection of craters during limited testing.

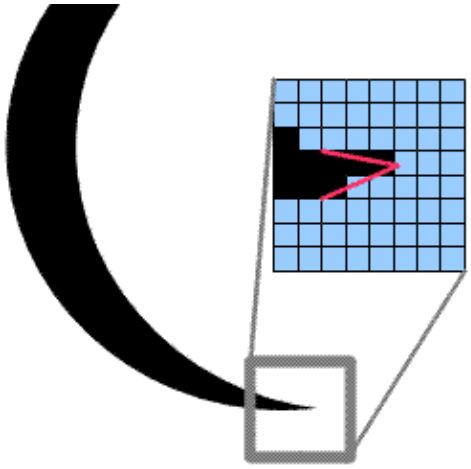

Figure 2: Two vectors are formed between closely neighbouring pixels. The angle between vectors is expected to be acute at extremal point on crater shadows.

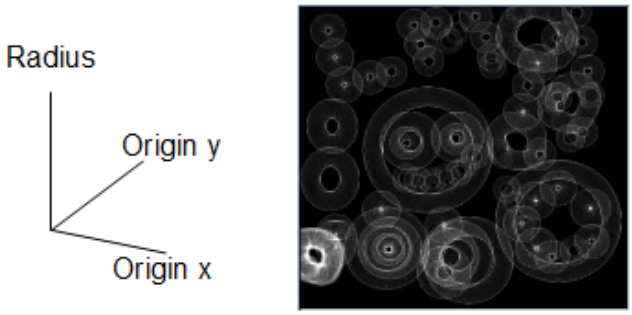

Figure 3: Left: the parameter space for a circle. Right: slice through the parameter space at a given radius. The bright spots show the accumulation of edges consistent with a circular interpretation with the given radius at the corresponding location in an image.

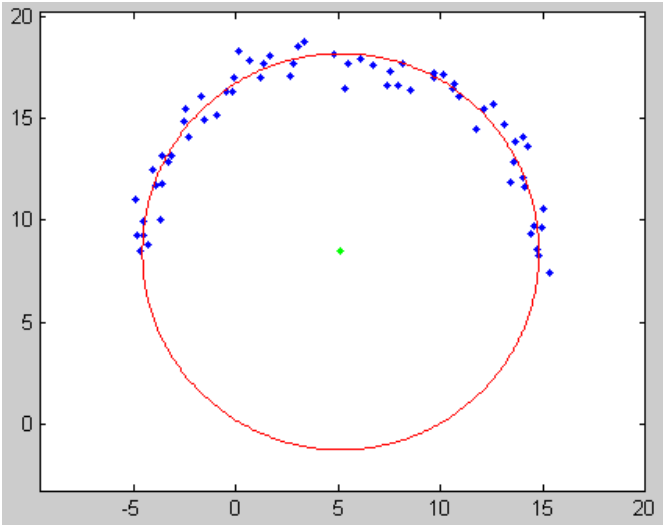

Figure 4: Circular conic section fitted to an arc of points. Various goodness of fit measurements include minimising the sum of squared residuals, median of squared residuals, mean residual, maximum residual and others.

3.1.2 Circular and elliptical edge detection methods

Numerous algorithms have proposed the use of a Hough transform [10] [11] to detect circular or elliptical arcs as a means of detecting craters [12] [13] [14] [19] [21]. The parameter space of a circle equation is partitioned into a number of discrete bins. An edge detector is employed to identify potential crater rim edges, combinations of which are then compared to possible circle parameters. Bins in the parameter space consistent with a circle interpretation of the edges are incremented. After all edges have been analysed the bins with the highest count indicate the most likely circles. This process is illustrated in Figure 3. Single arcs can be detected using this method, allowing for the possibility of partially occluded craters to be detected. Ellipses can be detected to account for perspective effects caused by different camera angles. Conic section fitting is an alternative approach for imposing similar interpretations of detected edges. A conic section is defined as the intersection of a plane with a cone as expressed by the following equation:

$$Ax^2 + Bxy + Cy^2 + Dx + Ey + F = 0 \tag{1}$$

where A, B and C are constants fixing the shape of a cone, D, E and F define the intersecting plane and x, y are Cartesian coordinates.

Various fitting algorithms are available which attempt to minimise the residuals between edges and a fitted circle, ellipse or conic [15] [16] as illustrated in Figure 4. The Hough transform is a reliable statistically valid approach to circular feature detection. The robust properties of the Hough transform are analysed in [17] where it is shown to be similar to regression methods such as least median of squares. The HT will successfully detect features in the presence of up 50% false or missing edges due to noise. Conic section fitting algorithms vary in their performance and stability. Both methods easily incorporate scale invariance.

Both the Hough transform and conic methods depend upon having reliable edges to work with. Surface images yield a great number of edges, many not associated with craters. Methods of preprocessing these edges include the thresholding of edge strengths [12] [19], use of morphological operations [21], selection of regions of interest using grey level co-occurrence matrices [18] and the removal of linear edges [19]. The removal of linearly aligned edges will upon occasion erroneously delete crater edges, a problem which becomes increasingly severe with interacting craters as depicted in Figure 5. Problems with such interacting craters are acknowledged by [18]. Other methods of circle fitting have been employed including genetic algorithms [21] and radial consistency [20].

Another complication of applying edge based methods is the illumination of craters which can produce multiple arcs between shadow transitions as illustrated in Figure 6. Some algorithms appear to make no explicate efforts to account for multiple arcs, or at least do not report their handling. Other algorithms make explicit assumptions about the nature of multiple arcs and attempt to extract the most appropriate arcs based upon the assumed shape of typical craters [18], raising the question of how the algorithm would perform in non-typical cases.

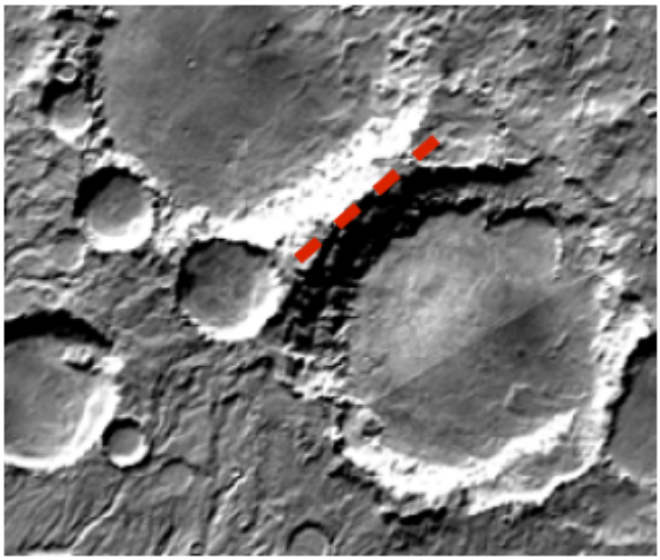

Figure 5: Misleading linear features become common in densely cratered images, just one problem associated in interacting craters.

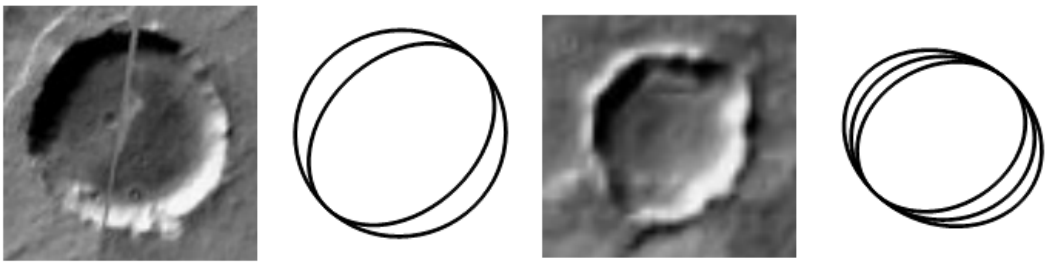

Figure 6: Multiple arcs are found within crater edge images. Left: most craters generate at least 4 arcs which outline the shadow and illuminated ridges. Right: some craters generate up to 6 arcs if the ridges are raised above the level of the surrounding surface leading to both inner and outer shadow and illuminated regions.

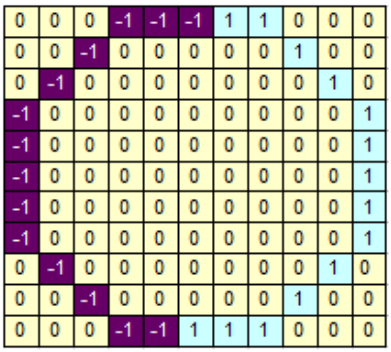

Figure 7: A convolution kernel can approximate a crater orientated in the direction of illumination. The negative regions respond to shadow and the positive regions respond to strong illumination.

3.1.3 Template and model matching methods

Annular convolution kernels have been used as matched filters which are sensitive to the direction of crater illumination [19]. One side of the kernel models the shadowed areas of a crater rim and the other models the illuminated side. Such a kernel is shown in Figure 7. The resultant images are searched for local maxima which can be labelled as potential craters. Craters of different sizes can be detected by either scaling the image or the kernel.

Naive convolution kernels only provide a rough model of a crater which are sensitive to many crater-like structures and channels giving many false positives. These models can be improved by taking a typical crater as a template and comparing it to all points in an image using normalised cross-correlation [19] [23]:

$$corr(i,j) = \frac{\sum_{k,l}^{n,m} f(i+k, j+l) g(k,l)}{\sqrt{\sum_{k,l}^{n,m} f(i+k, j+l)^2 \sum_{k,l}^{n,m} g(k,l)^2}} \tag{2}$$

where g is the n by m crater template and f the image under analysis.

A template crater is selected and compared using (2) at each point in an image under analysis. As with the matched filter, points of local maxima are identified as potential craters. Multiple templates can be used to match craters of different sizes and under different illumination conditions. This method tends to work best for small craters and cannot cope with interacting craters or craters with illumination vastly differing from the template. Unlike the shadow or edge methods, this method makes use of the full range of grey levels within an image making it potentially more discriminating. Other template matching methods are presented by [22].

The idea of a typical crater can be formalised by computing an average crater from a set of samples. The eigencraters borrowed from facial recognition literature [24] employ Principal Component Analysis to find the main forms of variation within crater images. All pixels within a sample crater form a feature vector. The eigenvectors of the co-variance matrix calculated from many sample vectors form the principal components. The components with the largest eigenvalues are retained as a lower dimensional description of a model crater. Eigencraters were used for false positive removal by [18], where the principal components of a putative crater are compared with the model and are rejected if they fall outside a threshold.

Eigencraters rely upon comparing craters in eigenvector space. Alternatively, an appearance model can be constructed which is compared in image space, as has been used in lip tracking [25]. A Point Distribution Model can pinpoint locations around a crater rim, allowing for elliptical or irregular craters to be modelled. PCA is used to identify the main variations in crater rim shape. One pixel wide cross-sectional profiles across the selected points are concatenated to form an overall grey level profile around the crater rim. PCA again is used to identify the main variations within the grey levels. A model crater can be reconstructed from the mean shape and profile which can be fitted with a cost function to detect craters in unseen images.

3.2 Drainage networks and channel-like structures

Algorithms for extracting extended channel-like structures must overcome some common obstacles including:

- large variability in channel length, shape and path

- branching and interaction between channels
- connecting channel fragments into complete networks

Far less focus has been given to the detection of such features on Mars in comparison to craters.

Methods to automatically extract drainage networks on Mars have focused on Digital Elevation Maps (DEM) generated by Mars Orbiter Laser Altimeter (MOLA) data [26] [27]. Such data is not affected by illumination conditions. These methods apply virtual rain over an area to model how a surface will drain. Each pixel within a DEM is assigned a pointer to the lowest of its neighbours. The pointer chains hierarchically coalesce forming accumulations of water theoretically corresponding to actual drainage channels. To ensure no localised minima unduly affects drainage a flooding stage identifies pits. The elevation of these pits is raised to the level of the lowest pour point around their edges thus allowing a surface to be fully drained. These methods were first developed for terrestrial river basins [28].

Fault scarps indicative of tectonic action has been extracted from DEMs using wavelet edge analysis [29]. A one dimensional filter based on a cubic B-spline function was used over 5 pixels. Stacking in a north-south direction formed a 5x5 convolution kernel which was tapered using a Hanning window in the east-west direction. The rotation and scaling of either an image under analysis or the kernel itself allows for features to be detected at different orientations and scales. It is reported erroneous features, particularly crater ridges, are identified using this method requiring elimination. To link and localise features hysteresis thresholding is employed, followed by hole filling and thinning morphological operations aimed at producing single pixel wide skeletons. Small braches deemed to be spurious are pruned before measurements being taken such as length and density of faults. No test results are formally presented for this method.

4 General solution - texture analysis

Craters, channel structures and other surface features have their distinctive properties, a-priori knowledge of which may aid in their detection using individually designed algorithms. Alternatively, a general solution to the detection of all surface features may be found in texture analysis. Texture analysis focuses on the recognition, segmentation and synthesis of images based on repeated textural elements [30]. Such textural elements could include the ripples of sand found within dunes or craters dispersed over a plane. Issues involved in texture analysis include defining suitable texture descriptors, handling texture boundaries, and accounting for invariances such as illumination, rotation and scale. Numerous texture analysis methods have been proposed, a selection of which are reviewed below. Many methods focus on producing texture descriptions which can be fed into a classifier. Others take a more direct statistical approach which directly build probability densities from textured images.

4.1 Texture Energy and Eigenfilters

Attempts have been made to describe textures by analysing their fine-grain structure. The more complex a region's structure the more energetic it is considered to be. Laws [31] proposed the use of three simple filters to detect local grey level intensity, local edges and spots:

- (1, 2, 1) local grey level
- (-1, 0, 1) edge detection
- (-1, 2, -1) spot detection

These simple primitives can be combined to search for wider structures such as ripples and waves.

- (1, -4, 6, -4, 1) ripple detector
- (-1, 2, 0, -2, 1) wave detector

These basic filters can be transposed and combined using matrix multiplication to create 2 dimensional structure detectors. The response to these filters highlights the correlations between pixels within texture primitives on a trail and error basis. Many texture filters may be required to fully describe a texture element and these can not generally be known a-priori. Once found, a selection of discriminating filter outputs can be fed into a classifier.

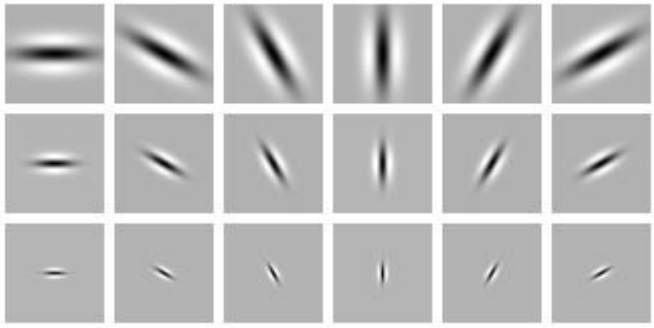

Figure 8: Series of Gabor filters for the detection of spacial frequency components at different orientations. Rows show increasing spacial frequencies downwards.

An alternative approach to finding the correlation between pixels in a textured region uses Eigenfilters proposed by Ade [32]. All possible pixel pairs within a small window are considered using a covariance matrix. The eigenvectors required to diagonalise this matrix expose the principal components. These components are analogous to the most discriminating texture energy filters previously selected by trail and error.

4.2 Fourier domain and autocorrelation

As texture elements often repeat periodically attempts have been made to categorise textures in the Fourier domain. This has been applied to the analysis of infrared satellite images [33]. Alternatives to crater counting have even been proposed which suggest Size Frequency Distributions could be replaced with Fourier analysis [34]. Problems with the Fourier analysis of textures include accounting for scaling, rotation and local deformations of texture elements which can have a large effect. But the largest problem faced by this approach is the global nature of a Fourier Transform as the 1 dimensional form below shows

$$FT(\xi) = \int_{-\infty}^{\infty} f(x) e^{-2\pi i x \xi} dx \tag{3}$$

where FT is the Fourier Transform returning the contribution of given frequency ξ within function f. Here, e is the base of the natural logarithm and i is the imaginary number. In a two dimensional case f could be an image. Clearly the integral required to calculate a spacial frequency's contribution to an image does not take into account any particular locality. Identifying the power spectrum for individual textures within a multi-textured image can only be done by applying a Fourier Transform on smaller image patches. This can be done by using a Windowed Fourier Transform which is more sensitive over a specified region, as defined by a window function in

$$WFT(\xi, t) = \int_{-\infty}^{\infty} f(x) w(x - t) e^{-2\pi i x \xi} dx \tag{4}$$

where w is the window function controlling f's contribution to the Fourier Transform and t positions the window along the x axis. If the w is a Gaussian the transform is known as a Gabor transform. Such transforms can be conveniently implemented as Gabor filters [36] [35] using convolution kernels such as those depicted in Figure 8. These filters can provide various channels of information such as granularity, periodicity and orientation at points in an image. By themselves these values do not constitute a full texture analysis system, as the probability of the resultant values conditional on particular textures must be addressed separately.

Autocorrelation is another possible approach to texture analysis. Measuring the granularity of textures has been attempted by counting how many pixels an image must be shifted before its autocorrelation drops below a threshold. Kaizer [37] proposed a threshold of 1/e whilst analysing aerial photographs. These measures of coarseness do not provide good information as to how isotropic or regularly distributed textural elements are, as in natural images these can vary in appearance and distribution over textured regions.

4.3 Wavelets

Wavelet based texture analysis is similar to a Windowed Fourier Transform and has been applied to Martian MOLA data [29]. A short-term wave segment, or wavelet, is any integrable function which can be scaled and

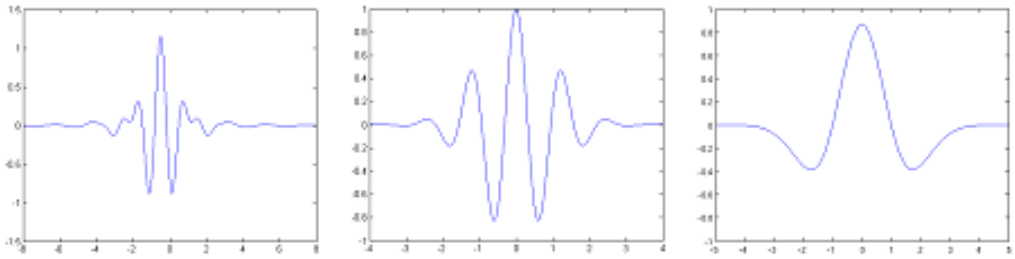

Figure 9: Example wavelet Mother Functions. Left: Meyer. Centre: Morlet. Right: Mexican Hat

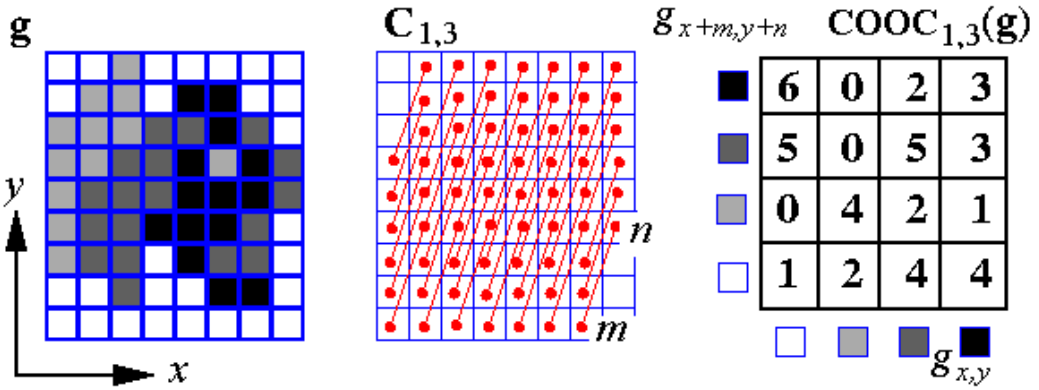

Figure 10: Left: patch of a textured image region containing grey levels g_{xy}. Centre: co-occurring grey levels a fixed vector apart. Right: Grey level co-occurrence matrix giving frequency of grey level co-occurrences.

shifted to model a signal. A series of such wavelets can describe the various oscillations in a textured region more efficiently than Gabor filters, as they have more expressive powers not being limited to sinusoid forms. Example wavelets include the Meyer, Morlet and Mexican Hat shown in Figure 9. The link between a wavelet transform and Windowed Fourier Transform is clear

$$\psi ab(x) = a^{-1/2}\psi(\frac{x-b}{a}) \tag{5}$$

$$WT(b,a) = \int_{-\infty}^{\infty} f(x)\psi^*_{ab}(x)dx \tag{6}$$

Where $\psi_{ab}(x)$ is the Mother Wavelet, a and b are scaling and shifting parameters repsectively, and $\psi^*_{ab}(x)$ is the complex conjugate of the Mother Wavelet. The Mother Wavelet defines the overall shape of a wave segment which can be used at different scales, which is especially convenient for texture analysis as texture elements are by their nature multi-scaled properties. Like Gabor filters, wavelets can be implemented as sets of 2 dimensional convolution kernels. A set of discriminating wavelets can then be used for classification.

4.4 Co-occurrence statistics

The most basic statistical method of differentiating between textures is to analyse grey levels. A grey level histogram and thresholding scheme can provide a probabilistic method of segmenting textures containing significantly different grey means. Whist this approach can be invariant to a great many parameters it is highly sensitive to illumination conditions, especially uneven illuminations [38]. It is also inappropriate for multi-modal textures.

A more sophisticated approach is to build probability densities of grey level co-occurrences, an early technique which has been applied to aerial photographs [39] [40] [41]. The co-occurrences of pixel grey levels in a textured region at a fixed vector from each other can be recorded within a 2 dimensional histogram. Once normalised these histograms yield probability mass functions against which other pixel pairs can be compared. Such a PMF is known as a Grey Level Covarience Matrix (GLCM). Their construction is illustrated in Figure 10.

There is little information within a single covariance matrix as texture elements cannot be fully modelled with just pixel pairs [42]. Also, problems occur when computing likelihoods based on co-occurrence statistics due to correlations between pixels. If pixel grey levels were uncorrelated it would be possible to relate the probability of finding particular grey levels, j and i, within a textured region given a set of pixels, Δ, at fix vectors, δ, apart using

$$P(j|\Delta) = P(j) \prod_{\delta}^{\Delta} \frac{P_\delta(i|j)}{P(j)P_\delta(i)} \tag{7}$$

where $P_\delta(i|j)$ is determined from a set of GLCMs. This however will not work in practice as the structure of textures ensures pixel values are correlated making the likelihood term invalid.

Algorithms designed to work with GLCM typically break away from the statistics, preferring to compute other values [39] such as:

$$Entropy = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} -P_\delta(i,j) \log(P_\delta(i,j)) \tag{8}$$

$$Contrast = \sum_{i=0}^{N_{g-1}} i^2 P_{x-y}(i) \tag{9}$$

$$Correlation = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} ijP_\delta(i,j) - \mu_x \mu_y}{\sigma_x \sigma_y} \tag{10}$$

where $P_\delta(i,j)$ is the joint probability of the co-occurrence of grey level i and j at fixed offset, N_g the total grey levels in an image, $P_{x-y}(i)$ the i^{th} entry in the distribution of absolute differences in co-occurring grey levels, μ_x, μ_y, σ_x, σ_y the mean and standard deviation of the two marginal probabilities of $P_\delta(i,j)$ respectively. Such values can then be fed into other classifiers [43].

To directly utilise grey level co-occurrence statistics it is necessary to overcome the correlation problem of combining pixel pairs as shown above where a product of multiple co-occurrence probabilities is computed. The most direct way to do this is to construct co-occurrence matrices for multiple pixels, i.e. constructing probability densities with one dimension per pixel [44]. This replaces the multiple GLCMs in the calculation of $P(j|\Delta)$ with a single matrix thereby modelling correlations and eliminating the need to compute products. For example, modelling the relationship between 3 correlated pixels would require a 3 dimensional matrix, 4 pixels a 4 dimensional matrix etc. Such density estimates will clearly grow rapidly with image patch size. It has been shown that such density estimates can differentiate between textures using maximum likelihood and EM, however the storage requirements become prohibitive when large numbers of correlations are being modelled [45]. Reducing the dimensionality of these matrices may make a larger scale algorithm tractable.

4.5 Markov Random Fields

Texture filter methods (Gabor, Texture Energy, Wavelets etc.) focus on the microstructure of textured regions. Attempts have been made to model the larger scale structure of textures using graphs which probabilistically connect regions of micro texture to neighbouring areas. The Markov Random Field is a popular choice, which assumes the micro-texture of a neighbouring region can be determined by learning the probability of transitions from training samples exhibiting the Markov property. This assumes the property of a textured sub-region only depends upon its immediate neighbours, i.e. the wider structure is a stochastic process which only depends upon the present (local) state. An example application involving the analysis of aerial photographs can be found in [46] where roads were segmented from urban scenes. Textures containing contextual dependencies beyond their immediate neighbours can be modelled using MRFs organised hierarchically into multiple resolution pyramids [47]. Research in this area includes the modelling and synthesis of textures, results of which can be seen in Figure 11. MRFs combine micro texture filter approaches with a statistical model with some similarities to GLCMs. Whereas a GLCM provides the probabilities of grey level transitions at fixed offsets, a MRF provides the probabilities of micro textures at adjacent nodes.

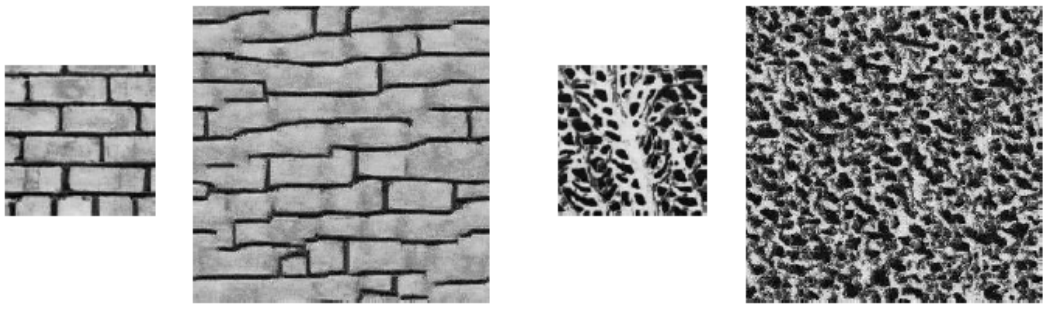

Figure 11: Examples texture synthesis using Markov Random Field models. Small images show sample textures along side their larger synthesised versions.

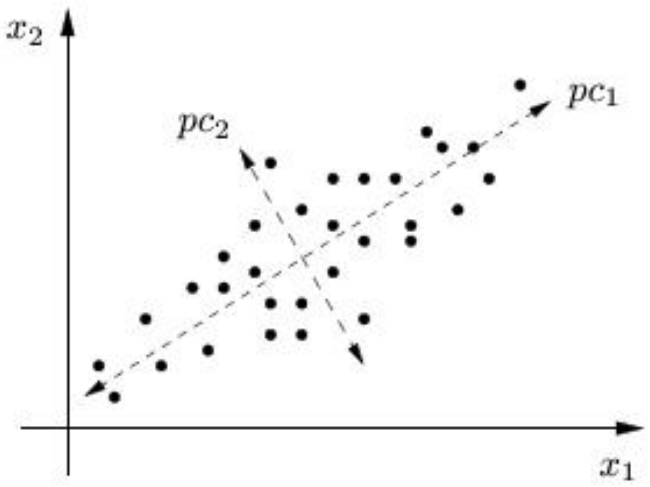

Figure 12: In 2 dimensions PCA can be seen as rotating data so new axis (principal components) account for the most variation in values. The covariance matrix of the new data space will be diagonal. Discarding the smaller principal components allows for lower dimensional representations at the loss of orthogonal information.

5 Dimensionality reduction

The high dimensionality of statistical texture analysis approaches was noted above. This section addresses the issue of how computational and storage requirements of such approaches can be made tractable. Dimensionality reduction techniques take high dimensional data and attempt to approximate the manifold upon which data points exist using a lower dimensional space. In order to appropriately model any manifold a method must take correct account of the underlying manifold's shape, nature of data distributions and errors on data points. Reduction techniques must minimise the residuals between actual data points and the approximated manifolds whilst not over fitting.

5.1 Principal Component Analysis

Principal Component Analysis (PCA) [48] computes the vectors within data space which account for the maximum amount of variance. PCA is equivalent to forming least squared hyperplane fits to data, as such errors on data points are assumed to be uniform and isotropic. PCA amounts to finding a projection of the original data onto an alternative coordinate system which yields a diagonal covariance matrix as illustrated in Figure 12.

$$S = \frac{1}{N}\sum_{i=1}^{N}(y_i - \bar{y})(y_i - \bar{y})^T \tag{11}$$

$$D = W^T S W \tag{12}$$

Where S is the original covariance matrix, y_i the data points and N the number of data points. Eigenvectors

of the data's covariance matrix are computed, the eigenvalues of which reflect the amount of variance in those directions. These vectors, w_i, are known as principal components and are used to diagonalise S forming D. The least significant principal components can be discarded thereby leaving a lower dimensional representation of the original data. PCA can be an effective method of reducing dimensionality as long as only linear relationships exist between dimensions and there is sufficient variance remaining in the retained principal components.

5.2 Kernel PCA

Kernel PCA [50] is an extension to PCA which attempts to model non-linear manifolds by mapping data points into a new space via a non-linear function (kernel).

$$S = \frac{1}{N}\sum_{i=1}^{N}\phi(y_i - \bar{y})\phi(y_i - \bar{y})^T \tag{13}$$

The kernel ϕ attempts to flatten out data so regular PCA can subsequently be applied. The kernel function itself contains an arbitrary number of degrees of freedom, the number of which will determine how accurately a manifold can be approximated. Too few degrees of freedom will lead to a rough approximation, whereas too many may cause over-fitting beyond the noise level and therefore poor generalisation to other datasets. The selection of kernel is therefore based on a trail and error.

5.3 Probabilistic PCA

Whereas PCA only attempts to extract the axes occupied by data, Probabilistic PCA is an extension which attempts to describe, in a probabilistic manner, a lower dimensional representation conditional upon its higher dimensional counterpart and a transformation matrix [51]. The parameters of the appropriate transformation can be determined using maximum likelihood on an iterative basis. This method assumes a high dimensional data vector, x_n, can be transformed into a lower dimensional approximation, y_n via a linear transformation, W:

$$y_n = Wx_n + \mu + \epsilon \tag{14}$$

where μ is the mean to which data points must be translated and ϵ is an isotropic normally distributed error with an expectation of zero. The transformation W is an l by m matrix where m is the dimensionality of the reduced data vector and l the dimensionality of the original data vector. PPCA has a stronger requirement than the similar Factor Analysis [51] which allows for non-isotropic errors.

Assuming data is first centred to give zero mean and data is normally distributed upon its manifold, PPCA gives the probability of observing a data point as

$$p(y_n|x_n, W, \sigma) \propto N(y_n|Wx_n, \sigma^2 I) \tag{15}$$

Integrating over x gives the marginal distribution

$$y_n \propto N(y_n|0, C) \tag{16}$$

where $C = WW^T + \sigma^2 I$. Assuming all data points are independent, the likelihood of observing a lower dimensional dataset Y conditioned on the transformation parameters can be calculate and maximised to determine the parameters W and σ:

$$p(Y|W, \sigma) = \prod_{n=1}^{N} p(y_n|W, \sigma) \tag{17}$$

$$ML = \arg\max_{W,\sigma} p(Y|W, \sigma) \tag{18}$$

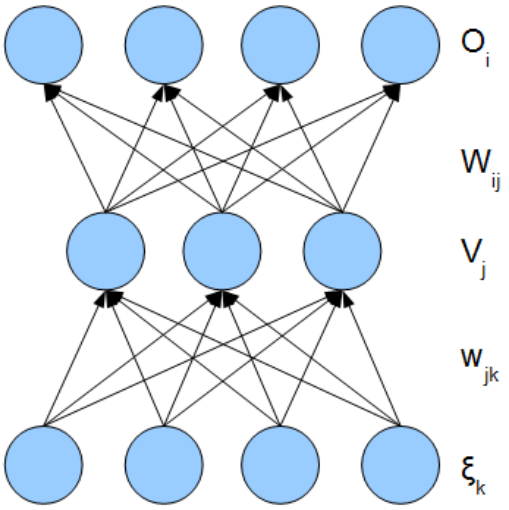

Figure 13: Feed forward neural network containing an input layer (bottom), hidden lay (centre) and output layer (top). Signals are propagate forward through the network as weighted sums at each node. Weight vectors W and w are adjusted iteratively in a process of back propagation.

5.4 Gaussian Process Latent Variable Models

Gaussian Process Latent Variable Models (GPLVMs) [52] generalise PPCA by replacing the linear transformation matrix W with an arbitrary function ϕ thereby combining the idea with that of Kernel PCA.

$$y_n = \phi(x_n) + \mu + \epsilon \tag{19}$$

Agreement between data and model is obtained by minimising the Kullback-Leibler divergence in the latent (lower dimensional) space.

5.5 Neural networks

A feed-forward back-propagation neural network trained to output a reproduction of its input can form a lower dimensional representation by feeding though a hidden layer of neurons [53]. Each node (neuron) contains input connections and output connections. The output is produced by passing a weighted sum of the inputs though a non-linear (typically sigmoid) activation function.

With reference to Figure 13, given an input pattern μ, hidden unit j recieves a weighted input

$$h_j^\mu = \sum_k w_{jk}\xi_k^\mu \tag{20}$$

yeilding an output

$$V_j^\mu = g(h_j^\mu) \tag{21}$$

where g is the non-linear activation function. The output unit i thus recieves

$$h_i^\mu = \sum_j W_{ij}V_j^\mu \tag{22}$$

yeilding a final output

$$O_i^\mu = g(\sum_j W_{ij}g(\sum_k w_{jk}\xi_k^\mu)) \tag{23}$$

A network is trained by systematically adjusting the weights between neurons using a back-propagation algorithm. Such an algorithm compares actual output with a desired output using a cost function

$$E = \frac{1}{2}\sum_{\mu,i}[\zeta_i^\mu - O_i^\mu]^2 \tag{24}$$

where ζ is the target output, which would be equal to the input ξ in a dimensionality reduction scenario. This cost function is appropriate for Gaussian distributed errors on the output data, and can be easily modified into a weighted least squares form if the measurement errors on the data are known. This choice can also be shown to approximate conditional probabilities of classification when training a network for '1 from C' coding [54].

Partial derivatives of the cost function with respect to each neuron's weight are assessed. Minimising the cost function becomes an iterative process of adjusting weights in the downward direction of the derivatives using gradient decent. For hidden-to-output and input-to-hidden weights the update functions are

$$\Delta W_{ij} = -\eta\frac{\partial E}{\partial W_{ij}} \tag{25}$$

$$\Delta w_{jk} = -\eta\frac{\partial E}{\partial w_{jk}} \tag{26}$$

where η is the learning rate which determines the speed of decent.

If the output is trained to reproduce the input and there are fewer nodes in the hidden layer, the hidden layer acts as a lower dimensional representation of the data. If the hidden layer, V_j, can be trained to maintain a metric space then it becomes a usable manifold approximation. A metric space will preserve euclidean distances between data points. The ratio of distances between any points in a lower dimensional space will therefore match corresponding ratios in the original space making linear interpolations between points meaningful and equivalent in either representation.

Problems associated with neural networks include the inordinate amount of time they can take to train. Techniques such as Boosting and Support Vector Machines were invented in order to circumvent the issues associated with training. However, these approaches loose the link to estimation of conditional probability. Methods which seek to regain this link using an additional mapping stage amount to using a low complexity neural network [49]. The possibility of training into local cost function minima rather than finding the optimal solution and the possibility of over training. An over trained network has learnt not only the underlying structure but also the noise found in training data. As noise will differ between datasets the network becomes unable to generalise. Identifying when training should stop is problematic. One solution can be to simulate noise on the sample data during training so no stable patterns of noise are seen during back-propagation [55].

6 Discussion

The image analysis methods presented above must be evaluated in the context of a scientific planetary analysis application. The following sections consider the minimum requirements of a practical working system and critically assesses the potential solutions.

6.1 Science requirements

For measurements made by an automated system to be scientifically valid they must, as a bare minimum:

- be quantitatively representative of the underlying truth;
- provide confidence intervals, i.e. error bars;
- and be repeatable

The ultimate output of an automated system will be a series of plots, such as SFDs, which can be compared against expected theoretical models. For the output to be useful it should quantify underlying features in inputted data, appropriately accounting for any systematic errors such as false positive and negative detections. Assuming classification errors are inevitable any analysis must make appropriate corrections to summary estimates. This can be done either by a separate analysis of efficiency and background (i.e. Monte Carlo), or as an integral part of the measurement. In this latter case, a practical system must be capable of providing honest estimates of the

probability any given feature detected is indeed what its purported to be. For example, to account for possible false positive crater detections an ambiguous feature should contribute to a plot proportionally to its probability of being a crater. If it is 50% certain a feature is a crater then the associated SFD bin should be incremented by 0.5. If 10 such ambiguous features exist only half will probably be actual craters therefore the quantitative assessment of the number of craters will be inline with the ground truth, i.e. 5.

Despite best efforts to correct plots there will still be uncertainties which must be made explicit via error bars. If such a confidence level cannot be assigned to a surface's age for example, it cannot be usefully compared to another. The interpretation of an SFD indicating one surface to be 100MA old and other to be 80MA old will differ greatly if the associated errors were 1% or 20%. With a 1% error one could be reasonably assured the 100MA old surface is indeed the oldest. On the other hand, with the presence of a 20% error both surfaces could be the same age or even the inverse of what was originally suspected.

If the same patch of surface is dated twice under different illumination conditions the interpretation of the two ages should not contradict. When compared, measurements taken under ideal and unfavourable crater counting conditions should yield consistent, i.e. repeatable, ages - albeit with explicitly indicated larger errors in the unfavourable cases. Similar arguments apply to any measurement, whether it be counting craters or parametrising fluvial channels or any other surface features. When assessing the suitability of existing image analysis methods these scientific requirements must be kept in mind. Any process manipulating input data which does not provide confidence and repeatability cannot be considered as part of a robust automated planetary surface analysis system.

6.2 Evaluation

6.2.1 Crater detection methods

The Smirnov method of crater detection suffers a number of shortcomings. Firstly, there are at least four arbitrarily selected thresholds, each with no principle behind their selection other than subjective trail and improvement. Such an approach immediately excludes the autonomous operation of the algorithm on large and varied datasets, as the repeatability necessary for scientific use is completely absent. Further to this, the measurements of discontinuities take no account of possible image noise which is especially problematic as angles are measured between vectors constructed from closely neighbouring pixels. A single mislabelled shadow pixel could swing the discontinuity angle wildly making the system very sensitive to errors in grey levels. The false positive removal steps are also problematic, as the random selection of a third reference point for circle fitting will lead to inconsistent rejection behaviour on the same image, even with all other parameters held fixed. Finally, the 100% reported performance during testing is exceedingly misleading, as the algorithm was only tested on a grand total of five craters. To be pedantic, assuming Poisson counting errors, the error on this figure is close to 45%. The estimated 80% efficiency is unsubstantiated and arguably highly optimistic. The Smirnov method is based largely on traditional binary image processing techniques as found in early machine vision textbooks. Such techniques were originally developed for use on very limited hardware. With modern equipment and more advanced image analysis methods now available this method is an inferior choice for a practical system.

The various edge based methods of crater detection employing either a Hough transform or fitting algorithm benefit from being scale invariant and tolerant to partial occlusion. These are great strengths given the nature of cratered surfaces. Unfortunately, these methods suffer significant common limitations largely due to the preprocessing involved in generating reliable edge images. Edges correspond to sharp discontinuities in grey levels which manifest themselves as first derivative peaks. Such peaks are influenced by noise yet no algorithm reports any systematic way of discriminating between significant peaks (i.e. above the noise level) and those that are not. This leads to the selection of arbitrary thresholds which may differ between datasets. Assuming a generalised significance test can be constructed to overcome these issues there are still problems associated with edges belonging to features other than craters. The removal of linear features may appear sensible but there is no reported method of determining how significant curvature has to be before a set of edges is retained. Plus, interacting craters can form linearities as shown earlier. Selecting regions of interest based on co-occurrence statistics prior to edge detection could potentially provide useful probabilistic interpretations of crater rims. However, rather than working directly with probabilities, algorithms compute values such as contrast and homogeneity from the GLCMs with no quantitative justification as to their application. The lack of a probabilistic framework capable of quantifying the significance of detected features severely limits an approaches' ability to correct for misclassification. Indeed, no method presented reports any attempt to correct results.

Template and model matching methods of crater detection can be viewed from a statistical perspective. The difference between a template or model in comparison to an area of an image under analysis is quantifiable. The larger the residuals between model and image patch the less likely an image patch belongs to the same distribution

as the model. Assuming residuals are independent, normally distributed random variables they can be seen as belonging to a chi-square distribution with degrees of freedom equalling the number of pixels in the model. The chi statistic can be computed and compared to a chi-square distribution yielding the probability that an image patch is consistent with a crater interpretation. This statistic is computed by

$$\chi^2_{ij} = \sum_{k=1}^{m}\sum_{l=1}^{n} \frac{[f(i+k, j+l) - g(k,l)]^2}{\sigma_f^2} \tag{27}$$

where χ^2_{ij} is the chi statistic for each location within an image, f is the image under analysis, σ_f^2 is the variance on the image and g is an m by n pixel template crater. Expanding the squared term, the numerator becomes:

$$f(i+k, j+l)^2 + g(k,l)^2 - 2f(i+k, j+l)g(k,l) \tag{28}$$

The convolution method of crater detection can be assessed against this theory. As g and σ_f are constant the convolution of a simple crater template with an image

$$\sum_{k=1}^{m}\sum_{l=1}^{n} f(i+k, j+l)g(k,l) \tag{29}$$

is clearly monotonically related to the chi statistic making local maxima good candidates for craters. However, this does not directly lead to a true probability for use in measurement corrections, nor is the convolution kernel a true model crater. Also, the illumination across f will vary making comparisons between different maxima difficult to interpret. Overall, the simple convolution method does not meet the scientific criteria.

The template matching crater detection methods via normalised cross correlation as given by equation (2) (repeated below for ease of reference) attempts to overcome differences in illumination across f by dividing by a normalising value. It also improves upon the template by using a real crater as a model.

$$corr(i,j) = \frac{\sum_{k,l}^{n,m} f(i+k, j+l)g(k,l)}{\sqrt{\sum_{k,l}^{n,m} f(i+k, j+l)^2 \sum_{k,l}^{n,m} g(k,l)^2}}$$

Again, it can be seen this is monotonically related to the chi statistic but does not directly result in true probabilities. Plus, the selection of crater template is subjective and need not necessarily correspond to the typical appearance of craters. To compound this, it is suggested different templates are used for different sized and illuminated craters meaning g is no longer constant. This results in correlation values which are no longer directly comparable. This method may be an improvement on convolution, but still scientific criteria is not met.

6.2.2 Drainage networks and channel detection methods

Methods to detect drainage and channel structures can be quickly eliminated from consideration as they are all based on Digital Elevation Maps rather than optical images. The extraction of such features from optical images has presumably been neglected as DEM are easier to work with and are readily available for Mars. Available software originally designed for terrestrial DEMs has been applied, but these were constructed for active river systems. Their direct application to Martian datasets assume the elevations of drainage networks billions of years old have not changed. To extract paleo-drainage networks which have been modified by tectonic processes it will be necessary to work with optical images.

Assuming DEMs were available, there are still issues facing alternative methods of channel structure detection. The detection of fault scarps, whilst intuitive, does not provide the necessary probabilistic information required for measurement corrections. The hysteresis thresholding for linking requires arbitrarily selected thresholds. The morphological operations designed to fill gaps and remove spurs are adversely affected by the spatial proximity of other structures and are not accompanied by any estimate as to how efficient the linking is.

6.2.3 Texture analysis methods

The analysis of fine-grain textures is dominated by methods of filtering using either sets of explicitly or automatically generated convolution kernels. The multichannel output of such filters is subsequently used to either identify regions using standard classifiers (e.g. Support Vector Machines) or segment regions with common properties (e.g.

K-means clustering). These methods do not in general provide quantitative information regarding the confidence one has in the classification process vital to correct scientific measurements in the target application. Learning only a decision boundary and forcing a classification may be appropriate in other contexts, but for an application such as crater counting it is desirable to account for the possibility of features right down to the noise level, even if this means counting fractional detections where the chance of correct classification is less than 50%. The multichannel filtering approaches by themselves can only therefore constitute a partial solution.

Grey level co-occurrences between pixel pairs and Hidden Markov Models do provide a level of probabilistic information for closely neighbouring pixels or regions but struggle to capture the wider correlations found within all but the most fine-grain stochastic textures. The correlation problems with combining GLCM has been noted and it can be seen in Figure 11 that MRFs can struggle to fully capture the process underlying a texture's generation. The assumption textures posses the Markov property does not always hold true as the probability of region transitions may change across an image, perhaps accounting for the mistakes in the brick texture in 11.

GLCMs, like convolution kernel filters, produce multichannel information via quantities such as contrast, entropy etc. which need further processing via classifiers, again only forming a partial solution. Multidimensional co-occurrences could potentially provide direct probabilistic classification information, but at a huge storage cost.

6.2.4 Dimensionality reduction

Straight forward PCA will approximate a linear manifold and can be used locally in patches to approximate a non-linear manifold to any desired granularity. Error propagation of PCA clearly shows that a lower dimensional representation of data will be well behaved making this a possible candidate for a working application. On the other hand, Kernel PCA can only be effective if input data is noise free as the non-linearity of kernels transform errors on data points making them non-linear also. Once these non-linear errors propagate through to the covariance matrix there can be no guarantee a least squared hyperplane fit will approximate the original manifold properly.

Probabilistic PCA assumes data is normally distributed upon its manifold. If data is not normally distributed this method will weight data points inappropriately. In a worst case scenario, if data is bimodal PPCA may assume the central tendency is in the trough between modes - exactly where the data should not be. The iterative maximum likelihood determination for parameter estimation will therefore only peak at the correct location in specific restricted circumstances. This is especially problematic for texture analysis as [56] encourage bi-modal parametric descriptions of textures which are shown to be significantly more accurate than simple Gaussian alternatives and [57] provided impressive results using more expressive non-parametric methods.

Gaussian Process Latent Variable Models attempt to incorporate both probabilistic and non-linear manifold approximations. As such they suffer the shortcomings of Kernel PCA and PPCA simultaneously. Between them, the non-linearities introduced by the kernel and strict requirement data be normally distributed prevents dimensionality reduction to be achieved reliably (i.e. minimised residuals between actual data and approximated manifold) in general cases.

Neural networks may provide a better, albeit slower, manifold approximation by minimising the back-propagation cost function (24). By constructing the cost function in the data space, from the difference between the output and the input, we can be sure the error needed for construction of the weighted least squares cost function is constant. This is in contrast to kernel PCA which attempts to find an optimal fit within the lower dimensional context (latent space). Here the non-linear function used to manipulate the data density distribution onto a hyper-plane can have adverse effects on the weighting of data in the minimised cost function (i.e. initially homogenous independant random noise can be amplified to generate outliers), so causing over fitting. Also, unlike the maximum likelihood of PPCA or GPLVM, neural networks minimise the local projection of data onto the manifold, but do not impose any restrictions on how data is distributed across it. It is therefore a more general approach which may be suitable for multi-modal textures.

Neural networks do carry a large initial computational overhead. Training involves many iterations, not only to determine appropriate weights, but also to select an optimal number of hidden nodes. Plus, there is always the possibility of training into a local minima, therefore many repeat runs may be required from different random starting points before confidence can be placed in a final network configuration. The popularity of alternative suboptimal approaches are perhaps due to them being quicker to execute. Such approaches may be of use in applications where speed takes precedence over accuracy. However, scientific requirements dictate a method be used which allows consistent and quantitatively accurate results making neural networks a potential solution.

7 Conclusions

Billions of dollars have been invested in exploring our planetary neighbours providing researchers with enormous quantities of valuable image data. Unless a reliable, statistically based automated method of analysing this data is found a valuable resource will remain tragically underutilised. Currently published research into the development of automated feature extraction from images has mainly focused on crater counting. Existing algorithms for surface feature exaction are mainly ad-hoc and do not address scientific requirements such as properly accounting for measurement errors, correcting output in light of false positive and negative detections, and providing reproducibility of results. Part of the difficulty in analysing planetary images is the inherent complexity, arising from the vast scope of possible features and terrains. A general solution to analysing such complex scenes may be found in texture analysis. Published methods of texture analysis also suffer from a lack of statistical rigour, with no off-the-shelf solutions sufficient for the task in hand. A statistically valid approach is likely to involve probability density estimates spanning a great number of dimensions. To make a large scale solution computationally tractable some method of dimensionality reduction is likely necessary. Dimensionality reduction techniques exist that can attempt to approximate a manifold on which textures can be modelled. The highly variable nature of textures exclude the use of a-priori knowledge in the modelling of data manifolds and distributions rendering the application of many dimensionality reduction techniques inappropriate. Locally approximating textures with linear patches or employing the flexibility of neural networks could provide potential solutions.

8 References

1. Mottmann J, Origin of the Late Heavy Bombardment, Dept. of Physics, California State University, Icarus vol 31, 1977
2. Michael G, Neukum G, Planetary Surface Dating from Crater Size-Frequency Distribution Measurements, Freie University Berlin, 2010
3. Bouley S et al. Comparison of Different Crater Counting Methods Applicated to Parana Valles, Center for Earth and Planetary Studies, Smithsonian Institution, Washington DC, 40th Lunar and Planetary Science Conference, 2009
4. Hartmann W, Crater Chronology and the Evolution of Mars, Planetary Science Institute, Tucson AZ, 2001
5. Warner N et al. Ancient Equatorial Thermokarst Lakes in Ares Vallis as Evidence for Transient Warm Conditions on Mars, Dept. Earth Science and Engineering, Imperial College London, 2009
6. Tyagi A, Identifying Areas of Differential Uplift Using Steepness Index in the Alaknanda Basis, Physical Research Laboratory, Dept. of Geology, Garhwal University, Current Science vol 97, No 10. 2009
7. VanLaningham S et al. The Effects of Rock Uplift and Rock Resistance on River Morphology in Subduction Zone Forearc Oregon, College of Oceanic and Atmospheric Sciences, Oregon State University, Earth Surf. Process. Landforms vol 31, 2006
8. Williams R, Terrestrial Inverted Channels in Utah: Analogues for Investigation of Martian Sinuous Ridges, Planetary Science Institute, Tucson AZ, 2008
9. Smirnov A, Exploratory Study of Automated Crater Detection Algorithm, Dept. of Computer Science, University of Colorado, Internal Report, 2002
10. Hough P, Method and Means for Recognizing Complex Patterns, US Patent 3069654, 1962
11. Duda R, Use of the Hough Transformation to Detect Lines and Curves in Pictures, Technical Note 36, Artificial Intelligence Center, Comm. ACM Vol 15, No. 1 pp. 11-15, 1972
12. Stepinski T, Machine Detection of Martian Craters from Digital Topography, Lunar and Planetary Institute, Houston Texas, Lunar and Planetary Science Conf., 2006
13. Salamuniccar G et a., Estimation of False Detections for Evaluation of Crater Detection Algorithms, Faculty of Engineering and Computing, University of Zagreb Croatia, Lunar and Planetary Science Conf., 2006
14. Chapman C et al. Identification and Interpretation of Martian Crater and Climate History, Southwest Research Institute, San Antonio, Texas, 2003
15. Al-Sharadqah A, Chernov N, Error Analysis for Circle Fitting Algorithms, Dept. of Mathematics, University of Alabama, Electronic Journal of Statistics vol 3, ISSN 1935-7524, 2009
16. Agin G, Fitting Ellipses and General Second-Order Curves, The Robotics Institute, Carnegie-Mellon University, 1981
17. Goldenshluger A, Zeevi A, The Hough Transform Estimator, Haifa University, The Annals of Statistics. Vol 32 No 5,2004
18. Kim J, Muller J, Morley J, Quantitative Assessment of Automated Crater Detection on Mars, Dept. of Geomatic Engineering, University College London, 2004
19. Magee M et al. Automated Identification of Martian Craters Using Image Processing, South West Institute, San Antonio, Texas, Lunar and Planetary Science Conf. 2003
20. Earl J et al. Automatic Recognition of Crater-like Structures in Terrestrial and Planetary Images, LogicaCMG UK Ltd, Space and Defense Division, Lunar and Planetary Science Conf, 2005
21. Honda R et al. Data Mining System for Planetary Images: Crater Detection and Categorization, Dept. Information Science, Kochi University, Kochi Japan, 2000
22. Vinogradova T, Burly M, Mjolsness E, Training of a Crater Detection Algorithm for Mars Crater Imagery, California Institute of Technology CA, 2001
23. Krogli S O, et al. Automatic and Semi-automatic Detection of Possible Meteorite Impact Structures in the Fennoscandian Shield Using Pattern Recognition of Spatial Data, Dept. of Geoscience, University of Oslo, 2007
24. Belhumeur P N et al. Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection, Pattern Analysis and Machine Intelligence, IEEE Trans, Vol 19 No 7, 1997
25. Luettin J, Thacker NA, Speech Reading Using Probabilistic Models, Images Science and Biomedical Engineering Division, University of Manchester, Tina Memo No 1997005, 1997
26. Stepinski T, Extraction of Martian Valley Networks from Digital Topography, Lunar and Planetary Institute, Houston Texas, Journal of Geophysical Research, vol 109, 2004
27. Stepinski T, Properties of Martian Highlands Drainage From Themis Image and MOLA Topography, Lunar and Planetary Institute, Houtson Texas, Lunar and Planetary Science Conf. 2006
28. Tarboton D, The Analysis of River Basins and Channel Networks Using Digital Terrain Data, Sc. D Thesis, MIT, Cambridge MA, 1989

29. Vaz DA et al. Automatic Detection and Classification of Fault Scarps on MOLA data, Centro de Geofisica, Universidade de Coimbra, Lunar and Planetary Science Conf, 2006
30. Chen C et al, The Handbook of Pattern Recognition and Computer Vision, 2nd edition, World Scientific Publishing Co, Chapter 2.1 pp 207, 1998
31. Laws K, Textured Image Segmentation, Report 940, Image Processing Institute, University of Southern California, 1980
32. Ade F,Characterization of Textures by Eigenfilter, Signal Processing, Vol. 5, pp. 451-457, 1983
33. Riou R, Seyler R, Texture Analysis of Tropical Rain Forest Infrared Satellite Images, Photogrammetric Engineering and Remote Sensing, Vol 63, No 5, pp 515, 1997
34. Morita S et al. Approach to Crater Chronology with Fourier Transform of Digital Terrain Models, University of Aizu Japan, Lunar and Planetary Science Conf. 2010
35. Jain A, Unsupervised Texture Segmentation Using Gabor Filters, Dept. Computer Science, Michigan State University, Proc. IEEE, 1990
36. Dunn D, Higgins W, Optimal Gabor Filters for Texture Segmentation, IEEE Trans. Image Processing vol 4, 1995.
37. Kaizer H, A Quantification of Textures on Aerial Photographs, MS Thesis, Boston University, 1955
38. Jain R, Machine Vision, McGraw-Hill International Publishing, ISBN 0-07-113407-7, pp 29, 1995
39. Haralick RM, Land use Classification Using Texture Information in ERTSA MSS Imagery, Remote Sensing Laboratory, Space Technology Center, University of Kansas, 1973
40. Haralick RM, Statistical and Structural Approaches to Texture, Proc. IEEE vol 67, No. 5 pp 786, 1979
41. Rosenfeld A, Comparative Study of Texture Measures for Terrain Classification, IEEE Trans. System Man and Cybernetics vol 6, 1976
42. Bromiley P, Thacker NA, Texture Recognition using Hypothesis Tests Based on Cooccurrence Distributions, Images Science and Biomedical Engineering Division, University of Manchester, Unpublished Tina Memo, 2005
43. Hong-Choon Ong, Improved Image Texture Classification Using Grey Level Co-occurrence Probabilities with Support Vector Machines Post Processing, European Journal of Scientific Research, EuroJournals Publishing, Vol 36. No. 1 pp 56, 2009
44. Kovalev V, Petrou M, Multidimensional Cooccurrence Matrices for Object Recognition and Matching, GMIP, 1996
45. Tar P, Thacker N A, Quantification of the Texture Analysis Problem, Imaging Science and Biomedical Engineering Division, University of Manchester, Tina Memo No 2010-001, 2010
46. Reulke R, Lippok A, Markov Random Fields (MRF)-Based Texture Segmentation for Road Detection, International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol XXXVII part B3b, 2008
47. Wei Li-Yi, Deterministic Texture Analysis and Synthesis using Tree Structure Vector Quantization, Gates Computer Science Building, Stanford University, 2000
48. Jolliffe I T, Principle Component Analysis, New York Springer, University of Geneva, ISBN: 0-387-96269-7, 1986
49. A. Niculescu-Mizil, R. Caruana,. Obtaining Calibrated Probabilities from Boosting, Proceedings of the 21st conference on uncertainty in artificial intelligence. Corvallis: AUAI Press, 2005.
50. Scholkopf B et al. Kernel Principle Component Analysis, Max Planc Institute, Kybernetik, Spemannstr, 1998
51. Tipping M, Bishop C, Probabilistic Principal Component Analysis, Microsoft Research, Cambridge UK, 1999
52. Lawrence N, The Gaussian Process Latent Variable Model, Department of Computer Science, University of Sheffield, 2006
53. Durai S, Saro A, Image Compression with Back-Propagation Neural Network using Cumulative Distribution Function, Word Academy of Science Engineering and Technology 17, 2006
54. Richard M.D, Lippmann R.P, Neural Network Classifiers Estimate Bayesian a-posterioi Probabilities, Neural Computation, 3, 4, pp 461-483, 1991.
55. Hertz J, Krogh A, Palmer R, Introduction to the Theory of Neural Computation, Santa Fe Institute Studies in the Sciences of Complexity, Addison-Wesley Publishing Company, ISBN 0-201-51560-1, 1991
56. Heess N et al. Learning Generative Texture Models with Extended Fields-of-Experts, University of Edinburgh School of Informatics, 2009
57. Efros A, Texture Synthesis by Non-parametric Sampling, Computer Science Division, University of California, IEEE Computer Vision Conference, 1999