Review of feature boundary identification techniques for the characterisation of tessellated surfaces

Gavin D. MacAulay\textsuperscript{1,2}, Nicola Senin\textsuperscript{3}, Claudiu L. Giusca\textsuperscript{1}, Richard K. Leach\textsuperscript{1}, and Atanas Ivanov\textsuperscript{2}

\textsuperscript{1}Engineering Measurement Division, National Physical Laboratory, Teddington, UK
\textsuperscript{2}School of Engineering and Design, Brunel University, Uxbridge, UK
\textsuperscript{3}Department of Engineering, University of Perugia, Perugia, Italy

Abstract

Tessellated surfaces are structured surfaces where the dominant features are organised in periodic patterns. Tessellated surfaces are becoming increasingly popular for a wide variety of industrial uses. However, their uptake is hindered by a lack of suitable metrological techniques to categorise and verify their properties. Areal surface texture parameters, commonly used for characterisation of surface texture, may not provide relevant information for characterising the periodicity and the other functionally relevant geometric attributes of the pattern. A possible solution, recently proposed in the literature, is to directly retrieve the individual features of the pattern and to determine their relevant geometric and dimensional properties. To achieve such a characterisation it is necessary to segment the surface in order to extract the pattern units (tiles) and the individual features contained within, so that their properties can be assessed. This paper reviews a number of different techniques and compares their ability to accurately identify the tile features and their boundaries. This ability is important as it has a direct impact on the computation of dimensional and geometric properties of the features.

1 Introduction

In recent years structured surfaces, i.e. surfaces with a dominant deterministic features designed to produce a specific function \cite{1, 2}, have become popular in a wide variety of industrial applications. In particular, tessellated surfaces, i.e. structured surfaces featuring periodic patterns, have shown promising results in applications, such as, friction reduction, optical devices and control of hydrophobicity \cite{1, 3–5}. It is, therefore, important that suitable metrological tools are available to characterise such surfaces, in order to verify their production and correlate with functional performance. This review illustrates the main approaches that have been proposed in the literature to characterise such surfaces.

1.1 Structured and tessellated surfaces

Evans and Bryan \cite{1} provided one of the first modern definitions of structured surfaces as, “surfaces with a deterministic pattern of usually high aspect ratio geometric features designed to give a specific function”. Subsequent researchers have added slight refinements and further subdivisions \cite{2, 6}. Stout and Blunt \cite{6} further subdivided structured surfaces into directional and non-directional surfaces depending on whether the pattern has a dominant direction. Similarly, Jiang and Whitehouse \cite{2} further subdivide structured surfaces into additional subcategories. Tessellated surfaces are structured surfaces with a periodic pattern. The most common subclass of tessellated surface has tiles with translational symmetry (e.g. retro-reflector surfaces consisting of a Cartesian pattern of hexagonal prisms, or the dimpled surface of a golf ball). The other subclasses of tessellated surface are linear patterned surfaces consisting of a single
linear texture unit which is only repeated in only one direction (e.g. ribbed or grooved structures) and rotationally invariant patterned surfaces where periodicity is in a polar coordinate system rather than a Cartesian one (e.g. Fresnel lenses). Finally, structured surfaces which are not classified as tessellated (i.e. they do not have a periodic pattern) are defined as multi-patterned surfaces, the most famous example being MEMS devices. Multi-patterned surfaces are not considered directly in this review, which focuses only on tessellated surfaces. However, many of the approaches discussed could be adapted for the analysis of such surfaces.

As shown in figure 1, tessellated surfaces are made of topologically connected pattern units, or tiles. Each tile consists of the same nominal topography, repeated multiple times to form the periodic pattern. The tiles are occupied by the deterministic features which make up the tessellated surface (e.g. the dimple of a dimpled pattern, a polygonal prism in retro-reflectors or abrasive surfaces). These features may cover all or part of the pattern unit. When the functionally relevant feature does not occupy the entire tile, the space between adjacent features can be referred to as the background. The distinction between feature and background is somewhat arbitrary, since the background surface is generally still important to the function, but is often adopted when characterising tessellated surfaces as it generally resembles the original surface design specifications.

Figure 1: Pattern units/tiles (blue) and individual features (red) of two different surface models: a) A dimpled surface where the features only fill part of the tile and are surrounded by background surface. b) A pyramidal surface where the feature fills the entire surface and no background exists.

1.2 Taxonomy of approaches to the characterisation of tessellated surfaces

Tessellated surfaces are formed of micro or nano-scale tiles repeated many times, so that the resulting pattern covers an area orders of magnitude larger than the scales of the individual features. Function in tessellated surfaces is generally exploited by the combined interaction of numerous tiles [1, 3]. It is, therefore, natural that the final goal of a procedure for characterising a tessellated surface is the assessment of properties pertaining to the pattern as a whole, instead of being limited to the characterisation of a single tile or feature. Conceptually, this is consistent with the approach adopted in surface metrology for characterising conventional surface texture where the ultimate goal is the identification of texture parameters, synthetic descriptors that provide information pertaining to the whole topography [7]. However, although the final goal of the characterisation is the same (i.e. computing parameters), the approaches adopted in the literature to obtain such results for tessellated surfaces have been varied, sometimes departing significantly from those adopted for more conventional stochastic (i.e. non-structured) surfaces. While a detailed illustration of the main existing approaches will be provided in the following sections of the review, a preliminary taxonomy is now introduced with the help of figure 2. Regardless of the
characterisation method, it is assumed that surface topography data has been obtained by an areal topography instrument and is available as a height map/range image, i.e. a set of pixels arranged into a regular grid.

Two general categories of characterisation route can be considered. The first route considers tessellated surfaces in a similar manner to how stochastic surfaces would normally be considered, i.e. field parameters are calculated based on statistical properties of the entire measured area. These parameters can either be based on the standard areal surface texture field parameters, described in ISO 25178-2 [7, 8], or on other, custom parameters. The approaches belonging to this characterisation route consider all measured pixels similarly, with no differentiation, and therefore ignore the tile structure of tessellated surfaces.

The second route encompasses many different approaches, sharing a common trait in that they introduce differentiation between measured points. In these approaches a few key steps are always recognisable: the surface is partitioned so that the individual tiles and/or features are isolated, and parameters are computed starting from the tiles/features, in the attempt to generate synthetic descriptors which are more closely linked to the function of the surface. There are a variety of ways of performing partitioning/feature identification, and there is a variety of ways to encode the results through parameters. One of the most common approaches consists of using morphological segmentation to partition the surface, and then computing the related ISO feature parameters [7]. However, many different, non-ISO approaches have been explored to segment the surface into tiles and features, and to compute parameters that describe the partitioned surface. Some of these methods target the tiles, others the features within the tiles, in those situations where the differentiation can be made (i.e. the feature does not entirely occupy the tile). In either case, the parameters proposed in the literature generally try to capture either tile/feature-related properties (i.e. tile/feature shape and size), or lattice-related properties (i.e. distance

Figure 2: Overview of the classification taxonomy for structured surfaces
between/alignment of the tiles/features).

2 Height map preprocessing

Areal topography measuring instruments, generally used in surface metrology, tend to record data as a raster image with heights in a regular grid (a height map). When measuring tessellated surfaces there are several factors that must be considered before the data can be analysed; most of which also apply for conventional (stochastic) surfaces. These factors include the treatment of non-measured points and measurement artefacts, filtering and levelling.

2.1 Treatment of non-measured points

Areal instruments often fail to measure the entire surface accurately. Optical instruments, in particular, may fail to collect enough information to determine the height of a pixel, which may be marked as non-measured (also referred to as a void). Some consistent method is needed to treat non-measured points so that the data can be used. There are two general strategies that can be adopted to deal with non-measured points. Either they can be left as non-measured, in which case algorithms capable of excluding them are needed, or they can be replaced with a plausible height value, computed from interpolation/fitting of neighbours. The first approach would be preferable from a metrological standpoint, because it would not introduce additional error in the data. However, adapting data analysis and processing algorithms so that they can discriminate between valid and non-valid points (e.g. through masking solutions) is generally not straightforward and speed/time performance may suffer [9, 10]. Therefore, correction of non-measured points is often a simpler solution. There are many standard approaches that can be used, usually based on interpolation/fitting of the surrounding valid pixels to come up with a best fit value. Notable examples include: linear interpolation [9], median interpolation [11, 12], splines [13] and kriging [13].

The problem of treating non-measured points exists for the characterisation of any kind of surface. However, it is particularly felt for structured and tessellated surfaces, when considering characterisation approaches targeting individual surface features because an individual feature is composed of far fewer pixels than a complete image. Therefore, missing pixel values have a much larger importance.

2.2 Treatment of measurement artefacts

Areal instruments sometimes produce pixels whose value is almost entirely the result of measurement error, generally because of specific phenomena taking place at the interaction between the probe measurement technology and the surface. These pixels generate false topographic formations (e.g. peaks) which are commonly referred to as measurement artefacts. Often measurement artefacts are characterised by height values that significantly differ from their immediate surroundings, and thus can be recognized by outlier-detection techniques. A typical example are the bat-wing formations generated by coherence scanning interferometers when encountering a step [14].

Unlike non-measured points measurement artefacts are not detected by the instrument. Therefore, an algorithmic approach is required to identify the measurement artefacts before treating them. For the applications and types of measurement error where it is safe to identify measurement artefacts through the detection of outliers, then multiple outlier detection techniques derived from statistics can be applied. The standard approach consists of estimating the probability density function of the heights of the good pixels in the area being investigated, and then identifying those pixels whose heights can be classified as outliers with respect to such distribution. An example of this approach is Grubbs’ test [14] which marks the point furthest from the mean as an outlier until the remaining points agree with a students t-distribution with \((N - 2)\) degrees of freedom \((N\) is the number of remaining pixels in the dataset). Many other similar algorithms exist.

One issue with such approaches is that they rely on the underlying assumptions of the outlier identification techniques. For surfaces consisting of uni-modal, simple-shaped local height distributions (e.g. normal) identifying outliers is generally easy, the ideal case being a horizontal rough surface with no...
significant high-scale formations. On the other hand, even the simplest step, or the presence of an underlying non-flat form or long spatial wavelength components, complicates the distribution of heights and makes it less stable as the surface is scanned, thus making it more difficult to detect outliers. Unfortunately, tessellated surfaces often contain such components. The typical solution to this problem is to remove form and long wavelength components from the topography, which can be done by subtracting a smoothed version of the original surface from the original surface itself, the residual containing only the higher spatial frequencies. The smoothed topography can be obtained by polynomial fitting on the original, although this technique does not work well in presence of steps and other sharp discontinuities, or by using a moving median filter [11, 12]. Another recent approach by Le Goic et al. is to use discrete modal decomposition [15]. Once the residual, high-frequency surface is obtained, outliers can be searched on it. Once identified, outlier pixels can either be excluded or corrected to more plausible height values, as for non-measured points. The identification and treatment of measurement artefacts is a common problem in surface metrology, but becomes particularly relevant for structured and tessellated surfaces, especially for those approaches requiring the inspection and characterisation of individual pattern features. As for missing values, this has to do with the fact that individual features are made of far fewer pixels, and the effect of local error is much stronger than when computing conventional texture parameters on the whole surface.

2.3 Filtering

It is often necessary to filter a measured surface to reduce high frequency noise. For stochastic surfaces a Gaussian filter is commonly used, which has the advantage of a monotonic frequency response [7, 16] and therefore does not produce any high frequency ringing effects. However for structured surfaces Gaussian filters are often poorly suited as they can distort the position of edges in the image and can therefore distort the features of the surface [17, 18]. If the features are large compared to the size of the smoothing kernel used then this distortion will be relatively small. However, as the feature size decreases the distortion effect will become more and more significant. Therefore, it is necessary to consider other, edge preserving, filtering techniques.

One candidate for edge preserving is morphological filtering [11, 16]. This filter is based on applying morphological open and closing operators to the surface, with spherical or flat structuring element. While it does a much better job than the Gaussian filter in terms of preserving significant edges, the conventional algorithms are computationally intensive especially for large structuring elements. This issue can make the filter prohibitively slow to apply, especially for large images. Recent research [19] has investigated novel algorithms to apply morphological filters which show promise to achieve much faster computation.

Another possible filter, proposed for use on structured surfaces by Jiang and Whitehouse [2], is the anisotropic diffusion filter [20–22]. This filter results in efficient smoothing of low-gradient regions while at the same time, sharp edges are preserved.

2.4 Levelling

Just as for measurements of stochastic surfaces, it is necessary to level measurements of structured surfaces to provide a consistent reference plane. The conventional way to do this is by subtraction of a least-squares mean plane. However, when considering structured surfaces it is necessary to exclude the surface features from the least-squares fitting as otherwise they can skew the mean plane [9]. It is important to note that, the plane should still be subtracted from the entire surface to ensure a single reference plane for the entire surface. To fit a least-squares plane to the background surface implies that the features have been identified and their position is known, for example using one of the segmentation methods described in section 4. However, the segmentation techniques often rely on having well levelled images. Therefore, an iterative approach is recommended whereby the image is segmented, then levelled, then segmented again until a stable image is reached [??]. An additional issue is that selectively fitting to the background is not possible for tessellated surfaces where the features contain the entirety of the tiles, such as in figure 1(b). Levelling against the whole surface may still be unsatisfactory if all the surface pixels are used, considering that the tiles may be cut differently at the image borders, and this could
skew the least-squares mean plane. The authors are not aware of any techniques designed for levelling of such surfaces.

3 Overviews of characterisation methods

As mentioned in section 1.2 there are several approaches that can be taken to characterise tessellated surfaces. This section provides an overview of these approaches, explaining the pros and cons and providing examples of each.

3.1 Characterisation without partitioning the topography

This section discusses the characterisation methods aimed at tessellated surfaces which do not require the topography to be first partitioned in order to identify tiles and/or the functionally-relevant features located within them. With no partitioning, surface points (pixels) are all treated equally, and included in the computation of the parameters. The techniques belonging to this category can be divided into those adopting ISO parameters for describing the topography (ISO 25178-2 [7]) and those that do not, and are based on devising original parameters instead.

3.1.1 ISO field parameters

Some researchers have attempted to characterise structured surfaces using ISO field parameters. In particular, Podgornik and Sedaleck [23] have tried, with some degree of success, to find a correlation between the kurtosis and skewness of laser textured dimples and grooves, and their coefficient of friction under boundary and mixed lubrication conditions. They succeeded in showing correlation between negative skewness and positive kurtosis parameters and friction coefficient. Such research results demonstrate that standard surface texture parameters may be useful for characterisation of structured surfaces. However, it must be questioned whether similar relationships between field parameters and function could be derived for other, unstructured, surfaces. Podgornik and Sedaleck note that the relationship they obtained follows the same pattern as that observed for friction in conventionally machined (i.e. unstructured) surfaces. In other words, a correlation was found between the distribution of heights of a surface and friction, however no information could be obtained about how being structured affects the functional performance of a surface, as many different feature shapes and layouts may lead to the same height distribution, and thus to the same related field parameters.

Accordingly, while ISO field parameters are widely used in analysis of surface texture for stochastic surfaces, several researchers have questioned their suitability for the analysis of structured surfaces. Weckenmann and Hartmann [24] claim field and profile parameters assume the surface has a random topography. Such parameters are therefore poorly suited to the analysis of surface with micro and nano structures and fail to map the geometry of the micro-structures successfully onto the functional performance of the surface, so that many micro-structured components can only be verified by functional tests. Similarly, Blunt and Xiao [25] argue that conventional surface texture parameters are designed to determine statistical properties of the entire surface whereas for structured surfaces the primary interest is in the deviations from the nominal of the individual features.

3.1.2 Custom parameters

Other researchers have attempted to define new parameters (i.e. non-ISO) in order to obtain a stronger relationship between the parameter and with functional performance, or even with the actual geometric properties of the features forming the pattern. Of particular note Geringer, Arneke and Seewig [26] propose a method based on a 3D analogy of granulometry. By calculating the volume of the surface after opening and closing operations with a range of structuring element sizes, plots of volume against structuring element size can be created. In the plot, the structuring element size corresponding to the maximum rate of change of the volume curve was found related to the characteristic lateral size of the
dominant features in the pattern. Geringer et al. tested this approach on a laser textured surface and showed a good correlation between the calculated and actual lateral feature size.

While this method appears promising for the characterisation of the specific features tested, care must be taken when applying this method to other surfaces. All the surfaces considered by Geringer et al. consist of features that are approximately circular in shape. It is not clear whether if other shape features are considered the size will be detected correctly. In particular, it appears that if severely elongated features were considered then only the smallest dimension of the feature would be detected correctly. Additionally, this approach is limited to considering the lateral size of the feature. While deeper features will give a stronger response, due to a greater change in volume, there is no way of directly determining feature depth or other properties which may be of interest.

As for ISO parameters, the main consideration here is that parameters referring to the whole surface topography may partially succeed in correlating with a surface functional property or specific shape/size property of the features on a structured surface. However, this correlation is indirect, potentially affected by many other factors, and hardly generalisable.

Another approach, suggested by Zeng et al. [27] uses the autocorrelation function to find the lattice properties of a tessellated surface by finding the translation vectors between peaks in the autocorrelation function (ACF). The translation vectors can be used to define parameters describing layout properties. Due to the regular nature of tessellated surfaces the ACF should consist of a number of regularly shaped, sharp peaks. By measuring the distance between such peaks the average lattice properties of the surface can be calculated. However, robustly determining significant peak position is not a trivial task.

3.2 Characterisation based on partitioning the topography

Many characterisation methods proposed in the literature imply a partitioning of the topography at some point. Partitioning, also referred to as segmentation, is either aimed at isolating the individual tiles, or at identifying and extracting the functionally relevant feature located within the tile (when the feature boundaries do not coincide with the tile boundaries). Once the partitioning is done, the final characterisation of the surface is based on computing parameters that capture the properties pertaining to the isolated tiles/features and to their spatial layout (lattice properties).

3.2.1 ISO morphological segmentation and feature parameters

ISO 25178-2 only recommends one method to partition the surface, morphological segmentation. In morphological segmentation the topography is partitioned into topologically connected hills or dales [11, 28, 29]. After segmentation is done, parameters can be computed that capture the properties of such hills or dales; these are known as ISO feature parameters [7]. For tessellated surfaces, the ISO approach to partitioning and computing parameters may be applicable, as long as a hill or dale segmentation can be appropriately configured to capture the tiles and/or the features contained within. This is a non-trivial problem, as illustrated in the work by Senin et al. [30]. In addition, even when each tile/feature is appropriately modelled by a subset of segments in the partitioning, it is still necessary to identify those feature parameters that better capture the properties of the tiles and of their layout. If successful, the combination of ISO morphological segmentation and ISO feature parameters allows the topography to be ultimately described in standardised and highly repeatable terms, thus favouring usability in terms of ease of information exchange in collaborative environments.

Hartmann and Loderer [31] have used morphological segmentation to identify the features in dimpled surface for friction reduction. Once features are identified their diameters and depths are calculated to produce distributions of diameter and depth for the entire surface.

Blunt, Xiao and Scott [25, 32] have considered the characterisation of laser textured hard disk drives with a tessellating pattern of bumps. The Laplacian of a Gaussian (LoG) operator is used before ISO morphological segmentation on the resulting image to ensure that the partitioned regions are representative of the real bumps. Parameters such as average diameter and depth of the bumps and average separation between bumps are then considered.
An entirely different strategy, still involving an ISO-compliant partitioning process, consists of applying a segmentation technique not directly to the original topography of the tessellated surface, but to an ancillary surface obtained from it. Jiang, Scott and Whitehouse [2, 33] have applied ISO morphological segmentation to the ACF of a tessellated surface. The partitioning leads to a robust identification of the general shape of the tile and therefore lattice parameters, because dale partitioning correctly identifies the shape of the region comprised within adjacent peaks in the ACF.

3.2.2 Other partitioning techniques and custom parameters

Many other methods have been used to partition the topography of a tessellated surface; they are discussed in detail in section 4. Accordingly, researchers have introduced various new parameters to describe the pattern units and their spatial layouts as obtained after the partitioning. For example, Kong et al. [34] used a height based threshold (see section 4.1) to identify the lenses in a micro-lens array. Various parameters were then used to characterise the micro-lens array, including the minimum, maximum and standard deviation of the roundness and lattice deviations in both a spatial and angular sense. Zhu et al. [35, 36] have considered an active contour (see section 4.3) based algorithm for segmentation of a range of surfaces, including etching silicon patterns, laser textured patterns and computer chip pins, and show that this method can accurately identify the surface features. Although, the calculation of relevant parameters is from these regions is not directly considered. Recently the authors of this paper have used thresholds on the gradient to partition laser textured silicon nitride surfaces into feature and background areas [37]. Using these partitions the mean and standard deviation of the diameter, volume and depth of a sample of the features were then calculated.

4 Segmentation methods

Segmentation, i.e. partitioning the surface into regions, can be used to identify tiles and features on a tessellated surface and then identify their boundaries. In the following section various segmentation methods will be examined and the application of such methods to feature/tile extraction in the specific case of tessellated surfaces will be discussed.

4.1 Thresholding

One of the most straightforward and widely used ways to segment surfaces in order to extract relevant features is to threshold the surface based on some local property of the pixels. Under this approach a threshold, \( T \), is set; all pixels with value less than \( T \) are marked as points in a feature and all points greater than \( T \) are marked as not belonging to the feature, or vice versa as appropriate for the surface. Many different local surface properties can be used for thresholding with differing effects as shown in figure 3.

Height of the pixel is probably the most straightforward and widely used threshold. Height thresholds provide an intuitive and computationally simple way to segment the surface. They are most effective when used to detect sharp edges, such as steps, and have been used extensively in the field of computer vision [38, 39]. Senin, Blunt, and Tolley [12] have recently devised a technique for the identification of thin foil laser targets for ion beam acceleration experiments, which at one step uses a height thresholding operation to discriminate between the target and background surfaces, as shown in figure 4. For tessellated surfaces, Kong et al. [34] have used a height threshold to identify the lenses and separate them from the background in a micro-lens array.

However, there are several issues with height thresholds which limit their applicability. Firstly it is difficult to select an appropriate threshold value. While several techniques exist to aid in selection of an appropriate threshold, including using the image histogram [11], Otsu’s method [40], k-means clustering [41–43] or using a percentage of a field parameter [9], these approaches rely on the height distribution being multi-modal and the feature being contained entirely within a well defined number of modes. Indeed, if this is not the case there is no algorithmic approach which can correctly identify the threshold
Figure 3: Effect of thresholding on example profile using different image properties. The principle is the same for areal data, profiles are shown here for ease of visualisation. In each case the red line shows the threshold value and feature areas are highlighted: a) surface height, b) gradient, c) local $S_q$ in $21 \times 21$ moving window.

[11]. Another concern is that height thresholds can produce many spurious features, such as those shown in figure 5, either due to noise in the measurement or the roughness of the background surface. While it may be possible to use filtering or other techniques to remove these falsely identified regions these approaches are unlikely to be perfect and must be tuned for a particular surface. A third serious concern when performing a height threshold is the effect of levelling. Height thresholds can perform poorly when dealing with unlevelled data because slopes can affect how pixels are marked. For example, if the background is tilted a threshold that successfully detects the feature is likely to also select a significant part of the background.

It is also possible to threshold on other local surface properties. One option is to use the value of a surface texture parameter as the threshold. From a metrological point of view surface texture is generally considered in terms of the areal surface texture parameters described in ISO 25178-2 [7] and is calculated as a single value for the entire surface or image. However, if these parameters are calculated in a local region around each pixel, rather than for the whole image, then a value can be attached to each pixel and thresholds can be set to segment the surface. Senin, Zilliotti, and Groppetti [43] have done this and succeeded in segmenting different textured squares and micro-indentations by using k-means clustering to determine thresholds on the local $S_q$ value.

Depending on how texture parameters are calculated, texture based thresholds may be less susceptible to levelling issues than height thresholds. Parameters involving local high-pass filtering or local levelling in their computation may end up being completely insensitive to global levelling of the topography. However, levelling effects are generally an issue to be considered in thresholding. There are also additional drawbacks to consider. Due to texture parameters being calculated on each pixel using information from a given amount of neighbouring pixels, a sharp transition in the topography (e.g. a step) may produce various results in terms of the transition of the local texture parameter, since at each position the
parameter would be the computed from an aggregation of pixels belonging to both sides of the step in the topography, with varying relative weights as the point moves across the transition itself. The simplest example of this would be the moving average, which results in a smoothed transition when applied to a sharp step. Most texture parameters computed locally over a region of neighbours would have similar behaviour, which must be taken into account.

Additionally, variations of the texture parameter value when computed over boundary pixels must be considered: at the image boundaries, fewer pixels are available to compute the local texture parameters, and this may affect the parameter value. Also choices on how to obtain the missing points (e.g. symmetric extrapolation, wraparound, constant value padding) affect the variation of the parameter, and thus the segmentation results.

Finally, texture thresholds suffer many of the same issues as height thresholds with regard to setting appropriate values for the threshold. Senin, Zilliotti, and Groppetti [43] used k-means clustering, with between two and six clusters to deal with this problem, which seems to have been quite effective for the surfaces they considered.

Another common option is to set the threshold based on the gradient of the surface. There are a number of ways to calculate the gradient of a surface. The Sobel filter is a popular method for range images, which convolves the surface with a pair of filter kernels to produce gradients in the $x$ and $y$ direction [11]. These maps can then be combined to produce the gradient magnitude, against which the threshold is applied. Setting such a threshold detects areas of high local slope, which generally includes the feature boundary and therefore separates the background and feature regions. Thresholding based on the gradient has similar pros and cons to texture based thresholds as they are closely related. In particular the choice of threshold is still an issue. As before automated methods can be used. However, on possible advantage of gradient based thresholds is that it is slightly easier to set the threshold manually as in many cases the expected gradient of the background is zero (or at least very small). However, in practise the choice of threshold still has a significant effect on the segmentation boundary. Additionally, the centre of features often have low gradient compared to the boundaries and therefore may not be marked as features, depending on how the threshold is set. As for texture based thresholds filling of enclosed regions can mitigate this issue. Recently, gradient based thresholds have been applied to the segmentation of laser textured silicon nitride disk [37] and successfully segmented the feature regions from the background.

Threshold based techniques are generally poorly suited to identifying pattern units (tiles) in tessellated surfaces as all the units will have similar properties and cover the entire surface. Depending on the specific surface it may be possible to set a threshold such that a region near the boundaries of the texture...
units are segmented from the rest of the surface. For example by setting a threshold on the height on the pyramidal surface in figure 1(b). However, this is still a non-ideal situation as the resulting boundary region will inevitability be thick and so accuracy in determining the tile boundaries will not be possible, unless some skeletonisation algorithm is applied [11].

4.2 Morphological segmentation

Morphological segmentation is an approach to segmentation based on the idea of morphological water-sheds proposed by Maxwell [44] in 1870. This idea is similar to the principle of the watershed of a river basin, in that the surface is segmented such that a drop of water placed at any point within a region flows down to the same point. The boundaries between regions are then the watershed lines, where a drop can flow in either direction. Reverse considerations lead to morphological segmentation into hills. Figure 6 shows a schematic of morphological segmentation for profile data. This can easily be generalised to areal data. Several different algorithms for efficient watershed segmentation have been developed [45–47]. These work either on the basis of immersion simulations (floodfill techniques) [46] where the image is flooded from beneath and watershed lines are defined where two flooded regions meet, or by considering how each point on the image flows to a minima (rainfall techniques) [45].

![Schematic of morphological segmentation on a profile](image)

One of the major issues with the watershed transform is that it tends to severely over-segment the image, as shown in figure 7 on a test surface of a laser-textured micro dimple. This oversegmentation is due to the fact that each local minima in the image will have its own watershed region. Due to noise in the image and roughness of the surface there will be many such regions which are unlikely to have any relevance to the surface function. Therefore, some method of region merging is required to merge the insignificant regions together into more significant regions, that ideally correspond to the structures on the surface. One of the simplest ways to merge regions is to smooth the image before segmentation using a Gaussian, or some other filter [48, 49]. This will reduce the number of minima in the image and therefore the number of watershed regions. However, such smoothing will also distort the image and cause the boundaries to move.

Another approach to deal with over-segmentation, often used in conventional image processing, is marker based segmentation [11, 50, 51]. Marker based segmentation relies on placing markers on the image which correspond to the significant features. The image is then morphologically deformed such that minima only exist at these markers. This reduces the number of watershed regions and helps them to agree with the significant features. However, marker based segmentation is not widely used in surface metrology because placement of markers requires significant a priori knowledge about the location of the features, either to place the markers manually or to control some algorithm to place them. This information is not generally available in conventional surface metrology but may be when considering structured surfaces.

A third approach to the problem is region merging techniques, of which several equivalent methods
exist [7, 52–54]. These rely on setting thresholds on properties of the watershed regions, often the depth and area. Watershed regions with values smaller than these parameters are then merged with neighbouring watersheds until all remaining regions are above the threshold. The problem with this approach is that the thresholds must be chosen heuristically as they will depend on both the desired segmentation and the surface. Therefore, there is no guarantee that a threshold which was successful for one surface will also be successful for another [30].

In surface metrology, morphological segmentation is primarily used as the method to segment a surface before determining feature parameters [7]. For this purpose region merging techniques are used to deal with over-segmentation. ISO 25178-2 [7] recommends using Wolf’s method [52] for region merging, although it points out that other methods, such as [53] are equivalent. Indeed, Scott [54] has shown that Wolf’s method provides a stable segmentation for determining surface features.

While apply morphological segmentation on height maps has been successful for many surfaces morphological segmentation can equally be applied to other surface maps, e.g. the gradient map [30, 55]. In many cases such approaches can be more successful in detecting the significant features. In particular morphological segmentation on the height map often produces unsatisfactory results when there is a large flat surface with a sharp step, as shown in figure 8(a), in such cases there is no sharp ridge and regions are merged such that the boundary lies a significant distance from the transition. On the other hand, segmentation on the gradient map, figure 8(b), is more satisfactory in such situations and places the transition in the centre of the transition, which is a ridge of high gradient.

Other maps can also be used to segment the surface. Blunt, Xiao and Scott [25, 32] have applied morphological segmentation to the surface after applying a LoG filter. This approach will detect the ridges of maximum curvature and has been successfully used to detect the dimpled structures of a laser textured hard disk drive.

In many ways morphological segmentation is ideally suited to detecting texture units as the segmentation approach can be seen as just looking for significant boundaries and determining regions based on the boundaries. Therefore, it is just a question of choosing a parameter such that the significant boundaries agree with the boundaries of the texture unit. This choice will depend on the surface in question. One method, proposed by Jiang, Scott and Whitehouse [33, 56], is to apply morphological segmentation to the autocorrelation function of the surface. Just as peaks are formed in the ACF due to the tile structure of the surface, the tile boundaries will lie on the valley lines of the ACF at minimal correlation. This approach has been shown to successfully identify the tiles for a range of surfaces.
4.3 Active contours

Another approach to segmentation is active contours [57–60]. Active contours start by placing a contour, \( C(s) \), on the surface and defining its energy, \( E(C) \). The contour is then evolved through space to try and minimise the energy. The final (minimum energy) contour then defines the boundary of the segmented object. Figure 9 shows the evolution of the algorithm on the test surface featuring a laser textured micro-dimple, with initial, intermediate and final contours shown. Clearly the definition of the contour energy will define where the boundary is placed. There are a wide range of energy functions that could be used. In image processing two of the most common are known as geometric active contours [58] and active contours without edges [60]. The geometric active contour [58], is based on an edge detection approach such that

\[
E(C) = \alpha \int_0^1 |C'(s)|^2 ds + \beta \int_0^1 |C''(s)|^2 ds - \lambda \int_0^1 |\nabla I(C(s))| ds
\]  

(1)

where \( C'(s) \) and \( C''(s) \) are the first and second derivatives along the contour, \( \alpha, \beta \) and \( \lambda \) are coefficients controlling the relative magnitude of each term, although commonly \( \beta = 0 \) is used, and \( \nabla I(C(s)) \) is the image gradient at that point. From (1) it is clear that, barring effects due to the contour shape, the energy will be minimised when the gradient is maximised.

An alternative approach, proposed by Chan and Vese [60], is active contours without edges. This method is more similar to the thresholding as section 4.1. It defines the contour energy

\[
E(C) = \mu.\text{Length}(C) + \nu.\text{Area(inside}(C))
\]

\[ + \lambda_1 \int_{\text{inside}(C)} |I(x,y) - c_1|^2 dxdy + \lambda_2 \int_{\text{outside}(C)} |I(x,y) - c_2|^2 dxdy \]

(2)

where the first two terms serve a similar purpose to the first two terms in (1) and similarly \( \nu = 0 \) is generally used. \( I(x,y) \) is the surface height at that point and \( c_1 \) and \( c_2 \) are the average height inside and outside the contour respectively. \( \lambda_1 \) and \( \lambda_2 \) are constants controlling the position of the segmentation, with \( \lambda_1 = \lambda_2 = 1 \) being the most common choice. This approach minimises the variance between the two groups and is therefore similar to k-means clustering [41] and Otsu’s method [40], which also try and minimise the inter-region variance.

One of the issues with active contours is the choice of initial contour. Ideally this should be placed close to the optimal contour to ensure swift convergence. If it is not, then the solution may take a long
time to converge, especially when there is a large background with little in the way of gradient for the contour to move down. In severe cases the algorithm may become stuck in a local minima and not converge to the optimal solution.

In surface metrology there has been some limited use of active contours for surface segmentation. In particular Zhu et al. [35, 36, 62] have used active contours with the geometric active contour algorithm for segmentation of various structured surfaces such as microprocessor chips and etched structures. In these cases they appear to be able to accurately segment the features, although no quantitative tests are made.

5 Feature identification and determination of geometric attributes

5.1 Feature identification

The aim of feature identification is to recognise a region of the topography as an instance of a class/category of geometry being searched for. When applied to tessellated surfaces, feature identification can be referred to the process of recognising regions of the topography as instances of the pattern unit (tile), or as instances of the functionally relevant features contained within the tile (if they do not coincide with it). Feature identification can be implemented as a subsequent step to segmentation, in which case the regions produced by segmentation are post-processed and analysed in order to recognise those to be associated with the target class, or without a segmentation step, typically through template-matching approaches.

5.1.1 Feature identification based on segmentation

This family of techniques first involves post-processing the regions resulting from segmentation (e.g. for removing the ones that do not comply to specific size and shape criteria, and/or for merging regions into more relevant formations, etc.) Common approaches are to remove regions below a certain size [37] or that are an abnormal shape (e.g. not round) [31]. Similarly some segmentation methods (e.g. gradient thresholds) may result in partitions that do not correspond to the feature being searched for, just to its boundary. To deal with this issue any enclosed region of background surround by a feature can be merged with the boundary to obtain a topologically connected (filled) region, as a good candidate for the subsequent feature identification step.

Another common issue is that regions intersecting the boundary of the image generally need to be excluded, as they may eventually result in features that are partially cropped out of the image, thus unsuitable for the determination of most geometric attributes. Finally, sometimes regions may need to be merged with each other before they can represent good candidates for feature identification. Whether this situation occurs is dependent on algorithm used, but is most common when using thresholding with multiple thresholds [43] or morphological segmentation [30]. In some cases this effect can be advantageous. For example, if the combined regions provide a more accurate description of the feature than could be achieved by segmenting the feature into a single region. In such situations these regions should be merged.
together to form a single feature. Suitable algorithms to perform such merging are often determined heuristically. A common approach is to mark the largest single region as the background and merge all other adjacent regions together. Post-processed regions are typically subjected to heuristic analysis in order to perform the final identification step. Typically identification is based on compliance to specific size, placement and form factor attributes. For example, in the test case involving laser textured micro dimples, segmented regions are identified as dimples based on their shape and size.

5.1.2 Feature identification based on template matching

Template matching is a feature identification approach that does not rely on segmentation. Rather than segmenting the surface and using the resulting regions to identify the feature, template matching starts with a geometric model (template) of the surface feature and moves this model over the surface to search for a best match with the local topography, as an indication of successful feature identification. Deviations between the identified features and the template can then be assessed to determine the feature’s properties.

Senin, Pini and Groppetti [63] describe template matching as consisting of three steps. First feature identification is performed. This finds candidate regions in the image where features matching the template can be found. Next comes feature extraction where the template is aligned with the feature in each candidate region and the region bounded by the template is extracted as a separate geometric entity. Finally the nominal (aligned template) and measured (extracted region) geometry are compared to determine deviations from the model. Each of these steps will now be considered in more detail.

Feature identification is one of the key steps in template matching. Some method must be used to compare the template and surface to find points where there is a good agreement between the two. This provides the rough location of the features which can then be accurately matched to the template. Many methods have been developed to perform this feature identification. Primarily these methods are based on finding a correlation between some property of the template and the surface.

Directly performing cross-correlation between the template and surface can be successful in some cases but is not ideal as the result is sensitive to the orientation of the template to the feature on the surface. Therefore a variety of other properties of the template have been investigated for locating the features. Senin, Pini and Groppetti [63] use a method based on the ring projection transform [64]. Jiang, Zhang and Scott [65] have also developed a similar technique called the structured region signature based on the point signature method [66]. The ring projection approach looks at the sum of around rings of different radii, whereas the structured region signature approach takes a single radius ring and looks for the best match at different locations.

Once the features have been identified coarse registration can be performed. This step aims to determine the rough alignment of the feature so that the fine alignment step can be performed successfully. If the alignment is already known for example if the template is rotationally invariant or is already determine by the feature identification step then coarse alignment is not necessary. Alternatively if the template consists of similar size area to the measured area, rather than just a single feature, then need for feature identification is lessened as course registration should cover all the significant features.

Again various methods exist. The most simple approach to coarse registration, as used by Senin, Pini and Groppetti [63] is to try a fixed number of orientations and determine which has the best fit between the template and the model. Yu et al. have developed an alternative approach based on matching the salient points of the template and surface [67]. This approach may be more complex but is well suited to cases where feature identification has not been explicitly performed and the location of the features is not yet well known.

Fine registration is used to create an accurate alignment between the template and surface. The most widely used algorithm for this step is the iterative closest point (ICP) method [68]. This algorithm computes the rigid transformation of the template which minimises the least squares sum of differences between the template and surface. However, being an iterative algorithm it is relatively computationally expensive and can easily become stuck in local minima. To avoid this issue good coarse registration is needed.
Once the surface and template are well aligned the difference of the two can be taken to determine form error in the surface model. This error could be assessed in a number of ways depending on the surface either using the sum of errors over the surface [63] or conventional dimensional tolerancing approaches such as tolerance zones [69].

5.2 Geometric attribute determination

The final step in determining feature parameters is to determine the relevant attributes of the individual features. The statistics of these attributes, computed from a sample of features collected while inspecting the measured area, are what constitutes the feature parameter. There are a vast array of different attributes that could be considered. It is therefore useful to consider attributes of the individual features and attributes of the pattern separately.

Attributes of the feature refer to properties of the individual features such as their shape and size. The method to determine the attribute will depend on the attribute in question. In general attributes of the features will either fit a shape to the boundary to determine its size (e.g. fitting a circle to determine feature radius) or will use some or all of the height data in the feature to determine the attribute (e.g. roughness or depth of the feature).

Attributes of the pattern, on the other hand, focus on the properties of the lattice which makes up the structured surface (e.g. the spacing and angle between features). To determine these attributes a consistent reference point is needed for each feature or tile to allow for consistent calculation of the parameters. For circular shaped features, such as the dimpled surface test case, one common approach, as applied by Kong et al. [34], is to fit a circle (or other appropriate shape) to the boundary of the feature and use the centre of the fitted shape as the centre of the feature. Other similar approaches could be used based on placing the centre at the centre of mass of the feature.

Once the relevant attributes have been computed feature parameters can be calculated by calculating the relevant statistics of the attribute distribution. Commonly used statistics are the mean, standard deviation and median. However, a wide range of other statistics could be considered, such as, mean of the five biggest/smallest values.

6 Discussion

Due to the wide range of characterisation techniques discussed in this review it is important to consider if there is an ideal method to characterise tessellated surfaces. On one hand, field parameters are well developed from characterisation of conventional surface texture. However, field parameters also have many limitations for characterising structured surfaces. As mentioned in section 3.1.1 field parameters generally do not account for the deterministic tile layout and shape, typical of tessellated surfaces. Therefore, they may be insensitive to changes in topography which change the functional performance of the surface. For this reason field parameters are often poorly suited to characterising tessellated surfaces. However, in situations where functional relationships with field parameters can be established or only verification is required field parameters may be suitable due to their relative simplicity and widespread understanding.

On the other hand, segmentation based methods produce feature parameters that are much more closely related to how such surfaces are designed. Such parameters allow for easier direct comparison with the specification and are more likely to correlate with functional performance. However, segmentation based techniques are generally more complicated with a larger number of control parameters to optimise. Within the range of segmentation based methods there are those that target individual features and those that target tiles. The two approaches have slightly different aims and are specialised for different purposes. Tile based methods have limited ability to determine properties of size or shape of the feature within the tile, as they do not separate the feature from the surrounding background, unless the feature is perfectly coincident with the tile that contains it. On the other hand, methods targeting individual features can still be used to determine lattice parameters of the surface, for example Kong et al. [34] determine spacing between the centroids of features after height thresholding.
For characterisation approaches relying on segmentation the choice of partitioning method should also be considered. Different segmentation methods will produce slightly different features and therefore will alter the resulting parameter values. Additionally the segmentation algorithms generally have at least one control parameter (e.g. threshold level, wolf pruning parameters) which can have a strong effect on the final segmentation results. It would therefore be of interest to consider how changing the control parameters affects the segmentation and attributes computed from it. The authors have recently presented an initial study of this effect [70] by considering the change in measured attributes when segmenting a particular surface with a variety of algorithms and varying the control parameters of each one. However, such an approach can only provide limited information as does not consider how this effect varies as the surface changes. Therefore it may be of interest to consider this effect in a more general way by considering a range of surfaces or by a modelling based approach.

Another outstanding issue is that quantifying the uncertainty in characterisation by segmentation based methods has not yet been attempted. One simple approach to calculating the uncertainty is via repeated measurement of the surface. However suitable experiments must be designed and performed to account for the wide range of factors that can vary between measurements. The disadvantages of such an approach are that it can be very time consuming to perform the large number of repeats required to account for all variations and the uncertainties achieved are not very general, they only apply to that particular measurement set-up.

An alternative approach would be to able to calculate uncertainties analytically based on knowledge of the uncertainties associated with the measurement instrument. The key steps here are to calculate the pixel-wise uncertainty in the measurement and then propagate it through the segmentation and characterisation processes. The easiest way to determine a pixel-wise measurement uncertainty is again by repeat measurement. This may somewhat defeat the point but it may be easier to develop a more generalisable model of measurement uncertainty in this way. Some work has been done determining analytical uncertainties associated with various segmentation methods, for example, De Santo et al. [71] considers the propagation of the image measurement uncertainty through the calculation of the Gaussian and gradient of the image and applies them to the calculation of lengths and areas of segmented images and their uncertainties. Additionally, De Santo et al [72] and Anchini et al [73] have calculated analytical uncertainties in various edge detection operators.

However, these approaches only consider specific segmentation methods and do not consider many of the pre and post processing steps, such as, smoothing and removal of small regions which are likely to be used in real processes. Many of these would be difficult to consider as they do not have simple analytical forms. A possible alternative to these approaches may be to consider a Monte Carlo approach [74]. If knowledge of the image measurement uncertainty is available this can be propagated through the algorithms in a statistical way without requiring detailed knowledge of how the uncertainty propagates through the algorithms. The drawback of such an approach is that it is very computationally expensive as numerous replicates must be calculated for each image.

Another issue, not so far discussed, is how to deal with the large lateral extent of patterned surfaces. Real structured surfaces will have large lateral extent, possibly up to several metres in size, and are likely to consist of many thousands of features. It is not always realistic to measure the entirety of such a surface with sufficient resolution to determine the properties of individual features. While using higher speed measurement instruments can mitigate this issue it in turn increases the measurement cost. Additionally as structured surfaces tend to work on the combined effect of multiple features knowledge of the mean and variance of the feature properties is often important. Therefore it may be possible to use standard statistical techniques to estimate these properties using only a sample of the surface features. However, in such a case it would be necessary to consider any spatial correlation between the structures which are likely to be created in the manufacturing process.

7 Summary

This review has considered a number of techniques to characterise tessellated surfaces. Field parameters are well developed from conventional surface metrology. However, they have limited ability to distinguish
between different tessellated surfaces and so are of limited use for these surfaces. On the other hand, tile and/or feature-oriented parameters, based on the identification of the tiles and/or the features within, show promise for use with tessellated surfaces as they can easily identify the dimensional parameters that are naturally specified during design of such surfaces. The segmentation and feature identification step is one of the key differences between these two approaches. Therefore, several different methods were examined and their strengths, weaknesses and applications were discussed. Additionally areas where further work is needed were discussed. These include attaching uncertainty values to the parameters associated with tessellated surfaces and how to deal with the sampling problem of measuring a surfaces consisting of many thousands of features with a large spatial extent.

References

19


