

Preprint. Final version appears as:

Newell, A., Morgan, R., Griffin, L. & Bull, P. (2010) Texture-based Estimation of Physical Characteristics of Sand Grains. *Digital Image Computing: Techniques and Applications*. DOI 10.1109/DICTA.2010.91

Texture-based Estimation of Physical Characteristics of Sand Grains

Newell, A., Morgan, R., Griffin, L. & Bull, P.

Communicating author: Ruth Morgan

Ruth.morgan@ucl.ac.uk

Tel: +44 20 3108 3037

Fax: +44 20 3108 3088

Texture-based Estimation of Physical Characteristics of Sand Grains

Andrew J. Newell
Department of Computer Science
University College London
a.newell@ucl.ac.uk

Ruth M. Morgan
Department of Security and Crime Science
Jill Dando Institute
University College London
ruth.morgan@ucl.ac.uk

Lewis D. Griffin
Department of Computer Science
University College London
l.griffin@cs.ucl.ac.uk

Peter A. Bull
Oxford University Centre for the Environment
University of Oxford
peter.bull@hertford.ox.ac.uk

Abstract

The common occurrence and transportability of quartz sand grains make them useful for forensic analysis, providing that grains can be accurately and consistently designated into prespecified types. Recent advances in the analysis of surface texture features found in scanning electron microscopy images of such grains have advanced this process. However, this requires expert knowledge that is not only time intensive, but also rare, meaning that automation is a highly attractive prospect if it were possible to achieve good levels of performance.

Basic Image Feature Columns (BIF Columns), which use local symmetry type to produce a highly invariant yet distinctive encoding, have shown leading performance in standard texture recognition tasks used in computer vision. However, the system has not previously been tested on a real world problem.

Here we demonstrate that the BIF Column system offers a simple yet effective solution to grain classification using surface texture. In a two class problem, where human level performance is expected to be perfect, the system classifies all but one grain from a sample of 88 correctly. In a harder task, where expert human performance is expected to be significantly less than perfect, our system achieves a correct classification rate of over 80%, with clear indications that performance can be improved if a larger dataset were available. Furthermore, very little tuning or adaptation has been necessary to achieve these results giving cause for optimism in the general applicability of this system to other texture classification problems in forensic analysis.

1. Introduction

Quartz sand grains have potential importance as a trace evidence for forensic investigations. This potential is based on two key features. First, the grains are ubiquitous in the environment and thus their occurrence in sources of evidence is common. Second, they have variable yet distinctive surface characteristics determined by their mode of formation and subsequent erosion, weathering and transportation[4]. The ability to reach exclusionary conclusions based on the provenance of quartz grains in forensic samples with the quartz grains identified in known samples is of value for forensic investigations.[14, 4]

The surface characteristics of grains are visible using a Scanning Electron Microscope (SEM), as shown in the images in Figures 2 and 3. With expert knowledge it is possible to use features from such images to place a grain within a classification tree [3], that can be used to designate grain types. However, such expert knowledge is rare and the manual identification of grains in this manner is time-intensive. Automatic identification of grains would therefore provide significant advantages in terms of making classification more widely available, as well as time efficiency and offering a standardisation of performance.

The earliest attempt[6] at a mathematical characterization of the physical characteristics of grains used Fourier methods to describe their shape. Since then other authors have proposed further methods based on shape[17, 13, 1], distribution of shape and size[16], and surface texture[20]. However, despite recent advances in texture recognition systems[18, 19, 2, 21, 5] there appear to be very few examples of these being applied to problems in the earth sciences. As far as we are aware, this is the first attempt to bring any of these techniques to grain analysis for applica-

tions in forensic analysis.

Basic Image Features (BIFs)[9, 7, 11, 10, 8] offer a simple yet effective solution to texture recognition, having been shown to produce state of the art results when using standard texture datasets in computer vision[5]. However, the system has never been tested on a real world problem. In this work we attempt to produce a simple yet effective method for quartz grain discrimination as well as provide a real world validation of the BIF system.

2. Texture Recognition

2.1. Encoding Texture

In the full BIF system, each pixel in an image is classified into one of seven types based upon the type of the approximate local symmetry. These approximate types are *flat*, *dark* and *light rotational*, *dark* and *light line*, *slope* and *saddle-like*. The classification is determined from the output of a bank of six derivative-of-Gaussian filters, one 0th order, two 1st order and three 2nd order. The algorithm has two tunable parameters. A filter scale parameter, a , and a threshold, ϵ , which is influential in deciding whether a locality should be classified as *flat*, or as one of the other six articulated symmetry types. Larger values of ϵ result in a greater proportion of an image being classified as *flat*. Although larger values of ϵ may be optimal for certain recognition tasks, previous work involving texture has indicated that the optimal value for ϵ is 0[5], so for this work the *flat* category is unused. The BIF calculation is given in Algorithm 1.

The resulting encoding for each pixel is invariant to rotation, reflection and affine scaling of the image intensity. By taking a histogram of the BIFs across an image and normalising we can also achieve an encoding scheme that is invariant to translation and the size of the image.

However, as there are only six BIF types in the scheme used for texture, such a histogram is unlikely to be specific enough for the discrimination of grain types. In order to overcome this, whilst still maintaining the invariance properties, we stack BIFs across different scales at each point in the image to create features called BIF Columns. [15]

Preliminary investigation suggested that the use of four scales would be optimal, and so we chose to use this number for this work. Our final encoding is now a normalised histogram with $6^4 (=1296)$ bins. The process is illustrated in Figure 1.

2.2. Recognising Texture

We represent images as 1296 bin histograms. To compare histograms we use the Bhattacharyya distance[12] as a metric. To classify an unseen image, we compute its histogram and use the metric to find the k Nearest Neighbouring (k NN) histograms from all histograms in a training set.

Algorithm 1 The BIF calculation

1. Measure filter responses $c_{i,j}$ to an (i,j) -order derivative-of-Gaussian filter, and from these calculate the scale normalised filter responses $s_{i,j} = a^{i+j}c_{i,j}$
 2. Compute $\sigma = s_{20} + s_{02}$, $\gamma = (s_{20} - s_{02})^2 + 4s_{21}^2$
 3. Classify $s_{i,j}$ according to the largest of:
 - (a) s_{00}
 - (b) s_{20}
 - (c) s_{02}
 - (d) s_{21}
 - (e) s_{11}
 - (f) s_{10}
 - (g) s_{12}
-

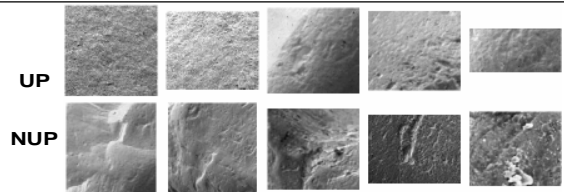


Figure 2. Example images from UP and NUP

If the number of the k having some label is above a threshold, we infer that the unseen image should have the label. Where possible the two parameters, k and the threshold, of our classifiers are determined using a validation set. However, in cases where the number of images per class is very small, these values may be set in advance.

3. Datasets

We chose two problems that are representative of the classification structure in forensic analysis, and assembled a dataset suitable for testing performance on them. The first of these was a two class problem, which involved detecting the presence of a geological feature called *Upturned Plates*.

Individual quartz grains were impacted against each other under aeolian conditions of known velocities under controlled laboratory conditions. Forty seven grains were subsequently imaged using scanning electron microscopy (SEM) and 266 images were taken of distinct areas of each grain which exhibited the *Upturned Plates* feature. This set of images was created using expert geological knowledge to ensure that the *Upturned Plates* features were present in each image and is referred to as the UP set. In order to create counterexamples to the UP set, 41 grains were selected by a geological expert from a library of quartz grains which displayed a range of alternative textures. From this set, 237 images were acquired using an SEM, none of which contained *Upturned Plates*. This set is referred to as the NUP set. Examples from UP and NUP are shown in Figure 2.

For the second problem the images in the UP set were further divided, by controlling the conditions of their formation, into six classes according to the *Energy Level of Formation* (ELF). The 266 images were labelled as 4mps,

IF Column gram Encoding

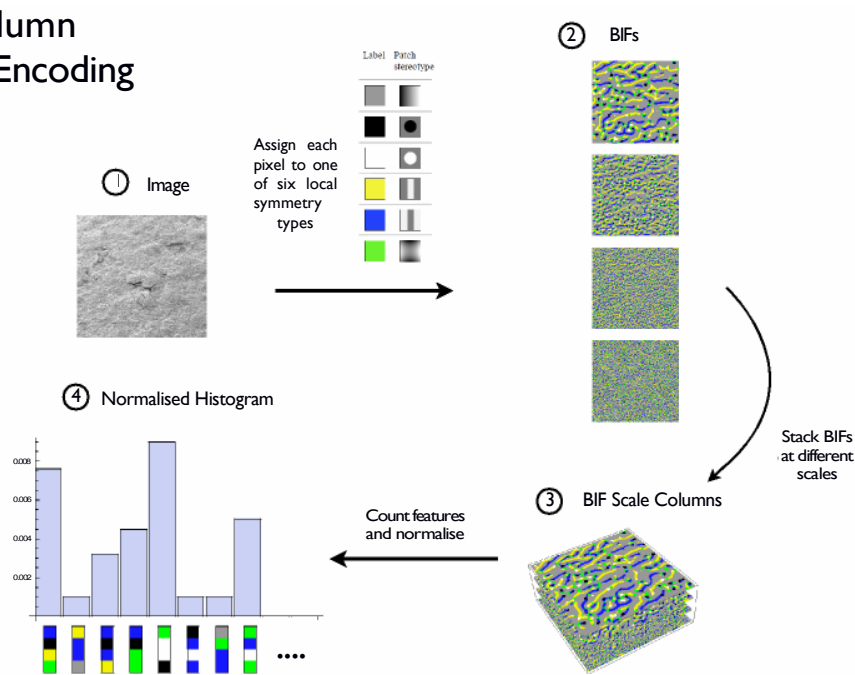


Figure 1. The stages in computation of a BIF Column histogram texture descriptor

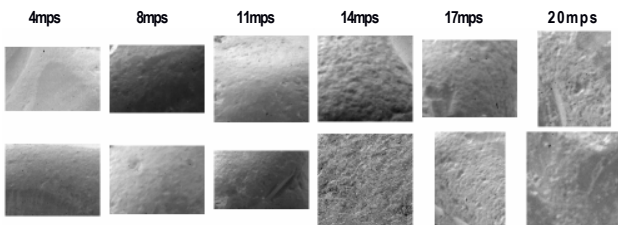


Figure 3. Example images from the set used in the *Energy Level of Formation* task

8mps, 11mps, 14mps, 17mps or 20mps, which relates to the wind speed to which the grains have been exposed. There was a minimum of 21 images from 6 grains per class. Examples from the different classes are shown in Figure 3.

Across the dataset, in addition to the textural differences that we sought to characterise, there were grosser types of difference that were due to imaging effects. Examples of global differences included variation in the overall lightness of the image, difference in the apparent focus of the image as well as difference in the image dimensions and overall area. Within images themselves there were variations in the lightness of individual regions, giving the appearance of shadows, although as the images came from the SEM these were not actual optical shadow effects. Each aspect of variation appeared within each class and there was no ap-

parent correlation between each non-textural aspect of variation and the class labels. We expected that the invariance properties of BIFs would mean that these types of variation would be ignored by our encoding as so not to lessen classification performance.

4. Upturned Plates

The images were encoded as 1296 bin histograms representing the frequency with which each BIF scale column occurred, as described in section 2.1. In order to make best use of the relatively small dataset without over-estimating performance by overlapping train and test sets, we used a nested leave-one-out method for both the test set and a validation set. To do this we first selected one grain as a test grain, with all images from this grain being extracted from the set. We then selected a single image from those left to act as a validation set, with images from the same grain being removed and the remainder used to build a *k*NN classifier. The validation image was then classified for different values of *k* and threshold. We then repeated this process for all images, except for those from the test grain, and found the optimal value of *k* and threshold. Using these optimal

Table 1. Performance for the *Upturned Plates* discrimination task

Scheme	Discrimination Performance Score
Without pooling	95.0%
With pooling	98.8%
Global Histogram	98.8%

parameter values we then classified the images from the test grain. This was then repeated with each image in the dataset in turn acting as the test grain.

As the problem being tackled was correct classification of a *grain*, and we had multiple *images* from each grain, we needed a scheme for making use of the multiple images. We evaluated three schemes for this to gain an understanding of the problem. First, we simply looked at how well the system performed with single images. Second, we used a simple pooling scheme where each image from a grain was classified individually and then a single choice was made for the grain by taking the more common classification across all images for that grain. Finally, we made an estimate of the global histogram encoding for the grain by taking the mean of the individual histograms for the different images from a grain. Results are shown in Table 1 for all three schemes, where they are quantified as the average of the classification rate for UP grains and for NUP grains.

For the three schemes the optimal values for the classifier parameters were determined individually for each different validation set and thus there were no single values that applied to the whole dataset. However, the median value of k for the classifier in the first two schemes was 15 and for the global histogram scheme it was 5.

5. Energy Level of Formation

Images were encoded as described previously. As the *Energy Level of Formation* task used only those images from the UP set, which were then divided into six classes, the number of images in each class were far fewer than in the *Upturned Plates* task. As a result, there were not enough images to provide a stable validation process to determine the optimal classifier parameters, so the value of k was set at 3 in advance and the threshold set at the midway point.

In order to determine the general discrimination power of the method in the ELF task, we first looked at the performance in discriminating each possible pair of classes. The same three schemes for grain classification were used as for *Upturned Plates*: based on a single image; based on pooling the classifications of individual images; and based on classifying the global histogram. Results are shown in Table 2.

From these results it was apparent that the system was performing poorly in discriminating between the 14mps, 17mps and 20mps classes, with results across all three scheme being consistent with chance level performance, as indicated in the grey boxes of Table 2. We therefore decided

to combine all images from these classes into one new class, labelled 14mps+.

We then looked at classification using the revised set of four classes. Using the same three schemes again, and a k NN classifier with k set at 3, results are reported as the mean performance per class. We report results in terms of exact classification (i.e. the correct one of the four classes is identified), classification to within one class, and classification to within two classes. These are shown in Table 3, with the associated confusion matrices given in Table 4.

Table 3. Performance for the Energy Level of Formation Task

Scheme	Exact	Within 1 Class	Within 2 Classes
Without pooling	69 %	90 %	100 %
With pooling	81 %	96 %	100 %
Global Histogram	78 %	92 %	96 %

As the number of images available for each class in the ELF task was relatively low, we also wanted to investigate whether we might expect performance to improve if more images were made available. To do this we looked at how performance changed when using a subset of grains from the 47 made available for the ELF task, using the pooling method as, out of the three, this had produced the best results.

Two grains were first randomly selected from each class to ensure the minimum size necessary to get results for the classification performance for each class. Then for each increment between this minimum of 8 grains and the full set of 47 grains, further grains were randomly selected and the classification performance determined for both exact and within one class classification. This process was repeated 400 times for each possible number of grains and the mean performance for each class over all trials is plotted in Figure 4. The mean performance over all classes is plotted as the black line, showing the overall performance as 81% when all grains are used, as previously shown in Table 3. This line has a steady gradient of 0.5% increase in performance for each grain added.

6. Summary and Conclusions

In the *Upturned Plates* task, where expert human level performance is expected to be perfect, the best performing scheme within the BIF system achieves 98.8%, which equates to classifying all but one grain correctly.

There is a clear improvement in the performance of the system when multiple images from a single grain are combined in some way. However, in this task there is no appar-

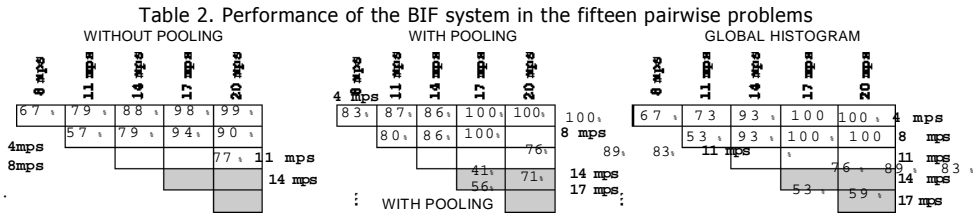


Table 4. The confusion matrices for the ELF task

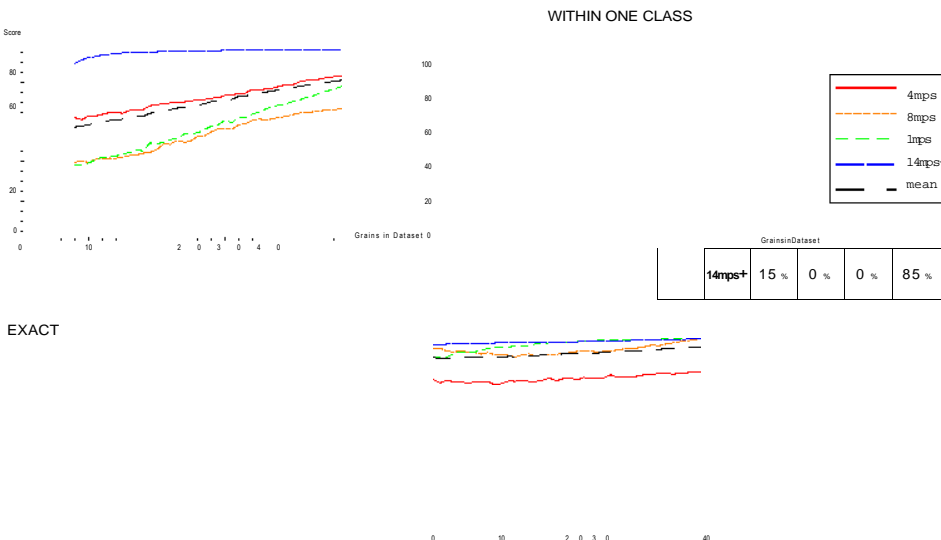
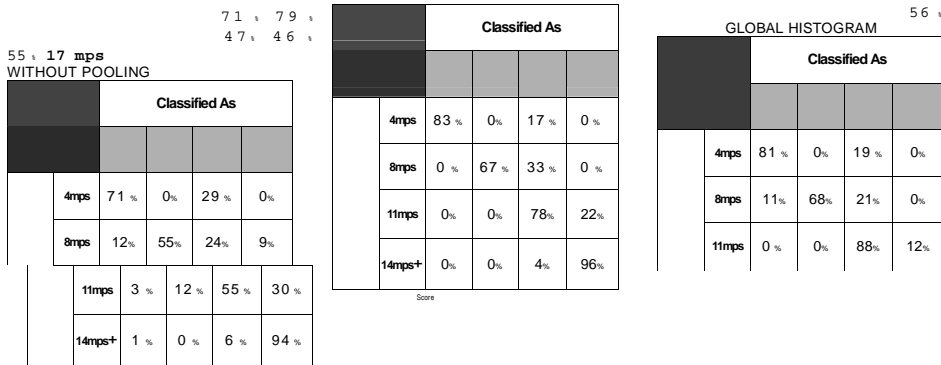


Figure 4. Performance with varying dataset sizes for each test class

ent advantage of one method of combining images over the other.

The *Energy Level of Formation* task represents a greater challenge. Expert human level performance, although not yet quantified, is expected to be significantly less than perfect, especially in discriminating the grains with higher ELFs. This is reflected in the results, where the BIF Column system is incapable of separating the 14mps, 17mps and 20mps classes from each other.

When these three classes are combined into one, discrimination performance between the four new classes is encouraging with the top performing scheme achieving a rate 81%

exact classification and 96% within an error of one class. As before, there is a clear advantage in combining the images from a single grain. In this task our results indicate a slight superiority of pooling image classifications over using the global histogram, but our dataset is not large enough to claim this is significant.

Our results indicate that improved performance could be expected with a larger training dataset, in particular for exact-class classification of grains in the three lower speed ELF classes. For the faster speed classes, and for within-one-class classification a dataset of 47 grains seems sufficient.

From the results as a whole it is evident that the BIF Column system can be used to provide an effective solution to the problem of grain discrimination using surface texture. We note that since the BIF column approach for encoding texture was applied without alteration or tuning from its previously presented formulation, its prospects for successful application to similar problems in forensic analysis are good.

References

- [1] E. T. Bowman, K. Soga, and T. Drummond. Particle shape characterization using Fourier analysis. *Geotechnique*, 51:545–554, 2001.
- [2] R. E. Broadhurst. Statistical estimation of histogram variation for texture classification. *Proceedings of the*

- fourth international workshop on texture analysis and synthesis. Beijing, China., pages 25–30, 2005.
- [3] P. Bull and R. Morgan. Sediment fingerprints: A forensic technique using quartz sand grains. *Science & Justice*, 46(2):107 – 124, 2006.
- [4] P. A. Bull, A. G. Parker, and M. R. M. The forensic analysis of soils and sediment taken from the cast of a footprint. *Forensic Science International*, 162:6–12, 2006.
- [5] M. Crosier and L. Griffin. Using basic image features for texture classification. *International Journal of Computer Vision*, 88(3):447–460, January 2010.
- [6] R. Ehrlich and B. Weinberg. An exact method for characterization of sand shape. *Journal of Sedimentary Petrology* 40, 40:205–212, 1970.
- [7] L. Griffin. Symmetries of 1 -d images. *Journal of Mathematical Imaging & Vision*, 31(2-3):157–164, July 2008.
- [8] L. Griffin, M. Lillholm, M. Crosier, and J. van Sande. Basic image features (BIFs) arising from approximate symmetry type. In *SSVM '09*, pages 343–355, 2009.
- [9] L. D. Griffin. The second order local-image-structure solid. *IEEE Trans. Pattern Anal. Mach. Intell.*, 29:1355–1366, 2007.
- [10] L. D. Griffin and M. Lillholm. Classifying local image symmetry using a co-localised family of linear filters. *Perception 37 ECVF Abstract Supplement*, page 122, 2008.
- [11] L. G. Griffin and M. Lillholm. Feature category systems for 2nd order local image structure induced by natural image statistics and otherwise. volume 6492, pages 1–11. SPIE, 2007.
- [12] T. Kailath. The divergence and bhattacharyya distance measures in signal selection. *IEEE Transactions on Communications Technology*, 15(1):52–60, 1967.
- [13] V. F. Leavers. Use of the two-dimensional radon transform to generate a taxonomy of shape for the characterization of abrasive powder particles. *IEEE Trans. Pattern Anal. Mach. Intell.*, 22(12): 1411–1423, 2000.
- [14] R. M. Morgan, P. Wiltshire, A. G. Parker, and P. A. Bull. The role of forensic geoscience in wildlife crime detection. *Forensic Science International*, 162:152–162, 2006.
- [15] A. J. Newell. A biologically plausible model for handwritten digit recognition. Master’s thesis, University College London, 2007.
- [16] N. Prakongkep, A. Suddhiprakarn, I. Kheoruenromne, and R. J. Gilkes. SEM image analysis for characterization of sand grains in Thai paddy soils. *Geoderma*, 156(1-2):20 –31, 2010.
- [17] M. C. Thomas, R. J. Wiltshire, and A. T. Williams. The use of Fourier descriptors in the classification of particle shape. *Sedimentology*, 42:635–645, 1996.
- [18] M. Varma and A. Zisserman. A statistical approach to texture classification from single images. *International Journal of Computer Vision*, 62(1):61–81, 2005.
- [19] M. Varma and A. Zisserman. A statistical approach to material classification using image patch exemplars. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(1 1):2032–2047, 2009.
- [20] A. T. Williams, R. J. Wiltshire, and M. C. Thomas. Sand grain analysis–image processing, textural algorithms and neural nets. *Computers & Geosciences*, 24(2):111 – 118, 1998.
- [21] J. Zhang, M. Marszalek, S. Lazebnik, and S. C. Local features and kernels for classification of texture and object categories: a comprehensive study. In *Conference on computer vision and pattern recognition workshop*, page 13, 2006.