"Virtual Sieve": Content-Based Image Retrieval in Automated Grading of Engineering Materials

Xiaoyu Qiao†, Fionn Murtagh†, Paul Walsh*, P.A.M. Basheer*, Adrian Long*, and Danny Crookes†

†*School of Computer Science*
Queen's University Belfast
Belfast BT7 1NN

**School of Civil Engineering*
Queen's University Belfast
Belfast BT7 1NN

E-mail: †x.qiao@qub.ac.uk *m.basheer@qub.ac.uk

Abstract — **The success of content-based image finding and retrieval is most marked when the user's requirements are very specific. An example of a specific application domain is the grading of engineering materials. In this paper we describe such an application in the area of civil engineering construction materials. We describe an innovative solution to automated vision-based grading of construction materials. From the methodology viewpoint, we show the advantages of using a resolution scale based approach for content characterization of mixtures of fine granularity material and large granularity "aggregate". In particular we use (i) multiscale entropy; and (ii) significant wavelet coefficients. Links with recent vision model perspectives are discussed.**

Keywords — **Machine vision, aggregate, construction, wavelet transform, entropy, information, image database**

I Introduction

a) The Civil Engineering Application

In terms of quality control the civil engineering crushed aggregate construction sector is relatively underdeveloped, with only simple manual tests being applied to the end product. From a cost-benefit perspective it is therefore essential that maximum economic value be obtained from the quarried stone, which will require wastage to be eliminated from each stage of the processing chain. The quality of the aggregate produced in terms of the consistency of its size and shape also has a major influence on the quality (particularly in relation to workability and durability) of the concrete and blacktop mixes subsequently produced.

Round or cubic shape aggregate particles have traditionally been considered the most suitable in relation to meeting the needs of industry, although it has also been suggested that bituminous mixes including non-cubic fractions can lead to better road pavement layer stability [1, 2].

The development of a rapid and efficient means for classifying aggregate size and shape could therefore enable the beneficial properties of an aggregate to be more fully exploited.

Aggregate sizing is carried out in the industrial context by passing the material over sieves or screens of particular sizes. Aggregate is a 3-dimensional material and as such need not necessarily meet the screen aperture size in all directions so as to pass through that screen. The British Standard specification (and American and other European specifications) suggest that any single size aggregate may contain a percentage of larger and smaller sizes, the magnitude of this percentage depending on the use to which the aggregate is to be put.

To monitor the range of size of aggregate particles produced from any particular screen, regular laboratory testing is carried out. This involves sampling the aggregate from either the moving conveyor belt or alternatively from the stockpile produced. A sieve analysis test is carried out to assess the range of particle sizes present in accordance with the specification. This test is time-consuming and therefore only a relatively small fraction (2 kg per 400–500 tons) of the aggregate produced is ever tested. The quality of the result also relies heavily on good sampling technique, which means that feedback to the quarry operators can be slow and in many cases unrepresentative.

Certain shape parameters are also specified for particular uses, the most common being Flakiness and the Elongation indices. These tests are also very labour intensive and time consuming, and are

carried out on an even more limited number of samples.

b) The Content-Based Image Retrieval Problem

An ability to measure the size and shape characteristics of an aggregate or mix of aggregate, ideally quickly, is therefore desirable to enable the most efficient use to be made of the aggregate and binder available.

This area of application is an ideal one for image content-based matching and retrieval, in support of automated grading. Compliance with mixture specification is tested by means of match against an image database of standard images.

For our work, the image data capture consists of an experimental environment which can be replicated in an operational setting: limitation of 3D effects and occlusion; use of diffuse homogeneous light; and avoidance of shadow.

For each class of aggregate mix, four separate samples were taken. Following each imaging, randomization was carried out on the aggregate mix. To provide a good sample in the case of each image, a subimage of dimensions 454×341 was extracted from a central region of each image. We took 50 images to represent each of the following aggregate classes, giving 600 images analyzed. A further set of 108 images (9 from each class) were used for testing. Classes: passing 6 mm sieve hole diameter; 30/40 mix; 50/10 mix; 10 mm; 14 mm; 28 mm drb; 40 mm; 28 mm dbc; 20 mm; 50-14 wc; 35 14 mm; and 510BC mm.

Figure 1 shows a representative image from the first of these classes.

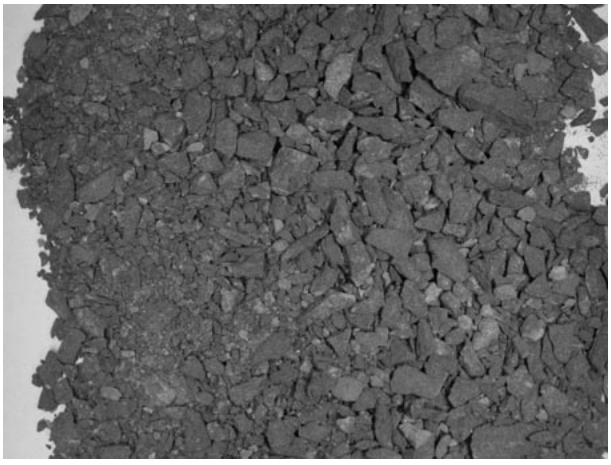

Fig. 1: Image from class 1.

Our objective is to create an image-based "virtual sieve" which, through image matching against an image database of standard images, will provide automated grading. We use an unsupervised feature selection and multiple discriminant analysis approach, to support nearest neighbour image querying.

c) The Vision Model Perspective

For robotic and industrial images, the objects to be detected and analyzed are usually solid bodies. The appropriate vision model for such images is therefore based on the detection of the surface edges (see e.g. [3]). However diffuse structures are characteristic of many other fields, including remote sensing, hydrodynamic flows, astronomy, and biological studies. Specific vision models are needed in each case. Such a model could be one for which the image is the sum of a slowly variable background with superimposed small-scale objects. In [4], a vision model is proposed where each pixel, at each level of spatial resolution, with a value significantly greater than the scale-based background is considered to belong to a real object. The same label is given to each significant pixel belonging to the same connected field, both spatially and in-scale (i.e. inter and intra wavelet band).

For the present work on civil engineering materials, we will use two vision models.

The first vision model will cater for fine grained material regions in an image. This image characterization is based on image entropy, determined by resolution scale (thereby catering for gradation in granularity size), and by spatial region (thereby catering for local subimage regions).

The second vision model will cater for coarse grained material regions in an image. It is based on significant wavelet coefficients at each resolution scale. Spatial (intra wavelet scale, and not inter scale) adjacency, only, is used to define object features. If it were necessary to analyze such objects, individually rather than globally, then inter scale analysis would be carried out.

II Feature Selection

a) Multiple Scale Entropy to Quantify Aggregate Granularity

For fine grained image characterization we carry out a "texture" analysis, and the wavelet transform is, by now, a traditional way to do this ([5, 6, 7, 8, 9]). Our approach avoids any system parameter related to window size; and the undecimated wavelet transform used helps to avoid object aliasing.

We used multiple scale image entropy to quantify aggregate granularity. Using 5 wavelet scales, from the B_3 spline à trous redundant wavelet transform, an entropy-per-scale was determined, and thus provided a 5-valued feature vector for each image. Background on this is provided in [10], where it was concluded that this approach to feature definition performed well for discrimination of

aggregate "textures".

We additionally used 5 wavelet scales, with the same wavelet transform method, used on the edge map, i.e. the image transformed with a Canny edge detector. In total, this provided 10 features per image.

A B_3 spline à trous wavelet transform gives the following decomposition of the original signal: $\{x_k \mid k = 1, 2, \ldots m\} = \{\sum_{j=1}^{l} w_{j,k} \mid k = 1, 2, \ldots m\}$. l is the number of scales, m is the number of samples in band (scale) m which is constant for this redundant transform. Scale l is the smooth or continuum scale, and all other scales consist of zero-mean (per scale) wavelet or detail coefficients. The value of l is set by the user (here, 6, implying 5 wavelet scales) and, given the dyadic property related to wavelet dilation, should be $< \log_2 m$. The feature set is defined from the resolution scale related decomposition as follows:

$$H = \{H_j \mid j = 1, 2, \ldots l-1\} =$$

$$\{\sum_{k=1}^{m} h(w_{j,k}) \mid j = 1, 2, \ldots l-1\} \quad (1)$$

with $h(w_{j,k}) = -\ln p(w_{j,k})$. The probability $p(w_{j,k})$ is the probability that the wavelet coefficient $w_{j,k}$ is due to noise. The smaller this probability, the more important will be the information relative to the wavelet coefficient. For Gaussian noise we have

$$h(w_{j,k}) = \frac{w_{j,k}^2}{2\sigma_j^2} + \text{Const.} \quad (2)$$

where σ_j is the noise at scale j. If we were using a (bi-) orthogonal wavelet transform with an L^2 normalization, we would have $\sigma_j = \sigma$ for all j, where σ is the noise standard deviation in the input data.

b) Larger Aggregate Characterization

Larger pieces of aggregate are less well modelled using the "texture" oriented approach described above. We used, in addition to the features used so far, a set of features designed to describe larger pieces of aggregate.

Firstly, this "larger aggregate pieces" description was multiresolution based. We used 5 wavelet scales of a B_3 spline à trous redundant wavelet transform. This transform does not favour any orientation, and being redundant avoids decimation-related effects. If one assumes a simple statistical model, such as a Gaussian one, it is straightforward to determine the Gaussian parameters in the wavelet planes. Note that the wavelet planes are linear combinations of the original image pixel values, and that any linear combination of Gaussian-distributed random variables yields a random variable which is Gaussian.

Feature 1 was the percentage of significant wavelet coefficients at each scale. Significance was determined from a 3σ threshold.

Feature 2 was the number of maxima at each scale.

Feature 3 was the number of structures (connected components of significant wavelet coefficients) at each scale.

Feature 4 was the size in pixels of the largest detected structure at each scale.

For 5 wavelet scales, using the aforementioned 4 features, this gave 20 features per image.

c) Multiple Discriminant Analysis

To facilitate assessment of discriminability between the classes in feature space, we used multiple discriminant analysis (also termed discriminant factor analysis, or the multi-class version of Fisher's linear discriminant analysis) [11, 12]. Discriminating axes are determined in this space, in such a way that optimal separation of the predefined groups is attained. As a linear discrimination method, we expect that such problems as training set size, and generalization, will be less pronounced than for a nonlinear method.

Consider the set of feature vectors, $i \in I$; they are characterized by a feature set, $j \in J$. A new orthogonal coordinate space is determined, such that the spread of class means in this new space is maximized, while the compactness of classes is restrained. Letting T be the total variance-covariance matrix of the n observations, and B be the between classes covariance matrix, we seek eigenvectors of the matrix product $T^{-1}B$ associated with non-increasing eigenvalues. It can be shown that multiple discriminant analysis is equivalent to a principal components analysis of centred vectors, i.e. the group means, in the T^{-1} or Mahalanobis metric.

Having the transformed feature vectors, i.e. their projection in the discriminant factor space, allows straightforward nearest mean assignment of vectors to the closest among the 5 groups used. In discriminant factor space, the (unweighted) Euclidean distance is used.

III Results

Figure 2 shows one example of the projected images in the principal discriminant plane, based on use of 5 classes. Classes 1, 2 and 5 are well separated. In the multidimensional space (inherent dimensionality 5 = minimum of numbers of: features less 1 due to centring; observation; and groups) the distinction between groups 3 and 4 is clearer. This figure used 250 images, characterized in 10-dimensional feature space. The first 5 features are the multiscale entropy ones described above. The second 5 features are multiscale entropies based

on a Canny edge transformed image. (See section IIa above.) For five successive classes of image, among the 250 images used, we had numbers of images misclassified as: 0, 0, 3, 7, 0).

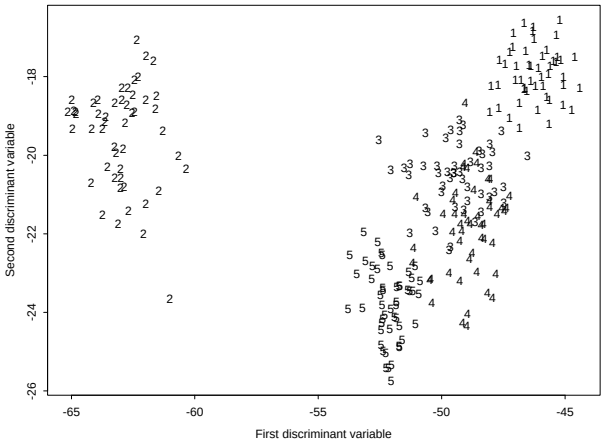

Fig. 2: Principal discriminant factor plane, with projections of images from groups 1 to 5.

Among the many experiments carried out, a few important points are as follows.

1. We studied many different feature sets. Most were multiscale-based. In particular we initially used multiscale entropy features of a Canny edge transformed images to provide information on larger pieces of aggregate. But we later bypassed these Canny edge-based features in favour of the multiscale significant structures analysis described above.

2. We examined the denoising of the image data prior to analysis. This was not found to be of benefit.

3. Rather than the nearest mean classifier used in the multiple discriminant analysis, we also investigated nearest neighbour discrimination approaches (1-NN, 3-NN); and a multilayer perceptron. However the linear approach was found to give very good results, and its operation was easily controlled and managed.

4. We worked on rebinned 454×340 images, obtained from the originally sized 2272×1704 images. We exhaustively tested the processing used on a battery of 600 training set images used additionally on the original 2272×1704 images. Days of compute time on a Sun Microsystems cluster gave us an approximate gain of 2% misclassification on an original 12%. We saw no justification in continuing to process the original images in this way.

5. We may note that as long as the misclassification in a class is shown to be less than 50%, then a majority class assignment based on a number of images is likely to increase the success rate.

Using 25 features (see sections II a and b: entropy values for 5 resolution scales; and 4 significant wavelet features for 5 resolution scales, all per image) and a set of 600 images, we found the following results: for 12 classes, numbers of images misclassified were: 0, 0, 6, 5, 3, 6, 15, 12, 9, 0, 2, 0. This gave an overall misclassification rate of 9.7%.

We then took 108 unseen iamges, and determined their classes based on the 600 image result. We found the following: classes 1, 2, 3, 5, 11, 12: all perfect. Class 4: 2 incorrect out of 9; class 6: 8 incorrect out of 9; class 7: 2 incorrect out of 9; class 8: 4 incorrect out of 9; class 9: 8 incorrect out of 9; class 10: 2 incorrect out of 9. Overall the misclassification rate was 24.1% incorrect. This result was appreciably improved with the following information. Classes 6 and 9 were two different classes but with mixture specification bands (British Standard BS 812 Part 103 1985) that were identical. Given an identical specification, no image-based virtual sieve can possibly separate these classes. In all test set cases, a majority class assignment leads to correct assignment.

In order to study the needs for training and test set cardinalities, assessments were carried out and are displayed graphically in Figures 3 and 4. For Figure 3, five randomly chosen test sets (and correspondingly randomly chosen training sets) yielded numbers of images misclassified as: 1, 0, 1, 1, 2, out of 30 in each case. Therefore we had an average 97% success on the test sets. In the case of the smaller training sets and bigger test sets exemplified in Figure 4, an average 95% success rate was obtained on the test sets.

Fig. 3: Principal discriminant factor plane, with projections of images from groups 1 to 6. The training set made use of 55 images from each group. The test set made use of 5 images from each group.

Fig. 4: Principal discriminant factor plane, with projections of images from groups 1 to 6. The training set made use of 50 images from each group. The test set made use of 10 images from each group.

Specification of mixes are within standard bands. We have now started to investigate necessary specification band coverage. Not surprisingly our initial results show that good coverage is needed. In other words, the feature space of training set examplars needs good coverage relative to the test set cases.

IV Conclusions

We have tested a new image content characterization approach, with excellent results on images of aggregates containing different object sizes and morphologies. Our algorithms are computationally inexpensive, and scalable. In our experimental evaluation, we have found these algorithms to be robust and stable.

From the point of view of operational use in the difficult conditions of the construction industry, we note that the algorithmic robustness and stability leaves just one area where care and attention will be required in practice: viz., the operational camera and lighting environment.

References

[1] P. Hobeda. Krossningens betydelse på stenkvalitet, starskilt med avseende på kornform. Literaturstudie Nr. 050001, Statnes vag-och Trafikinstitut, VTI, Linköping, Sweden, 1988.

[2] V. Reinhardt. Schlagfester Splitt 8–11mm oder stabiler Asphaltbeton 0–12mm. *Bitumen, Teere, Asphalte, Peche und verwandte Stoffe*, 1969. Nr. 11.

[3] H. Choi and R.G. Baraniuk. Multiscale image segmentation using wavelet-domain hidden Markov models. *IEEE Transactions on Image Processing*, 10:1309–1321, 2001.

[4] J.L. Starck, F. Murtagh, and A. Bijaoui. *Image Processing and Data Analysis: The Multiscale Approach*. Cambridge University Press, 1998.

[5] N. Fatemi-Ghomi. *Performance Measures for Wavelet-Based Segmentation Algorithms*. PhD thesis, Surrey University, 1997.

[6] S.G. Mallat. A theory of multiresolution signal decomposition: the wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11:674–693, 1989.

[7] S. Livens, P. Scheunders, G. Van de Wouwer, D. Van Dyck, H. Smets, J. Winkelmans, and W. Bogaerts. A texture analysis approach to corrosion image classification. *Microscopy, Microanalysis, Microstructures*, 7:1–10, 1996.

[8] M. Unser. Texture classification and segmentation using wavelet frames. *IEEE Transactions on Image Processing*, 4:1549–1560, 1995.

[9] P. Scheunders, S. Livens, G. Van de Wouwer, P. Vautrot, and D. Van Dyck. Wavelet-based texture analysis. *International Journal of Computer Science and Information Management*, 1(2):22–34, 1998.

[10] F. Murtagh, Xiaoyu Qiao, D. Crookes, P. Walsh, P.A.M. Basheer, and A. Long. Machine vision methods for the grading of crushed aggregate. *Machine Vision and Applications*, 2003. submitted.

[11] F. Murtagh and A Heck. *Multivariate Data Analysis*. Kluwer, 1987.

[12] J.M. Romeder. *Méthodes et Programmes d'Analyse Discriminante*. Dunod, 1973.