Machine Vision & Applications manuscript No.
(will be inserted by the editor)

A Machine Vision Approach to the Grading of Crushed Aggregate

Fionn Murtagh[1⋆], Xiaoyu Qiao[2], Danny Crookes[2], Paul Walsh[3], P.A. Muhammed Basheer[3], Adrian Long[3] and Jean-Luc Starck[4]

[1] Department of Computer Science, Royal Holloway, University of London, Egham, Surrey TW20 0EX, UK
[2] School of Computer Science, Queen's University Belfast, Belfast BT7 1NN, UK
[3] School of Civil Engineering, Queen's University Belfast, Belfast BT7 1NN, UK
[4] DAPNIA/SEI-SAP, CEA-Saclay, F-91191 Gif-sur-Yvette Cedex, France

Received: date / Revised version: date

Abstract The grading of crushed aggregate is carried out usually by sieving. We describe a new image-based approach to the automatic grading of such materials. The operational problem addressed is where the camera is located directly over a conveyor belt. Our approach characterizes the information content of each image, taking into account relative variation in the pixel data, and resolution scale. In feature space, we find very good class separation using a multidimensional linear classifier. The innovation in this work includes (i) introducing an effective image-based approach into this application area, and (ii) our supervised classification using wavelet entropy-based features.

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

1 Introduction

1.1 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 [8,16].

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

⋆ Corresponding author, F. Murtagh, e-mail fmurtagh@acm.org, tel. +44 1784 443429, fax +44 1784 439786

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 labor intensive and time consuming, and are carried out on an even more limited number of samples.

Systems based on individual objects (e.g. in quarrying) have been proposed [23,22]. The collective particle analysis proposed by us in this paper is less usual. The sieving metaphor is not to be taken literally in that small, fine particles would be obscured or pass quickly through sieve filters compared to larger, gross aggregate. Our virtual sieve work assumes instead that imaging is carried out on a production line conveyor belt with due attention paid to vibration (to avoid losing the small particles). As exemplified in the following sections, a layer of aggregate is imaged, with the camera position directly above the field of view, and with the layer of aggregate approximately covering the field of view. Uniform light is used to avoid shadows. Some obscuring of finer particles is unavoidable but we show experimentally that the overall system is robust and effective.

1.2 The Machine Vision 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 was extracted from a central region of each image. We took 50 images to represent each of the following 12 aggregate classes, giving 600 images analyzed. A further set of 108 images (9 from each class) was used for further testing. Classes were defined as passing: (1) 6 mm sieve hole diameter; (2) 30/14 mix; (3) 50/10 mix; (4) 10 mm; (5) 14 mm; (6) 28 mm drb; (7) 40 mm; (8) 28 mm dbc; (9) 20 mm; (10) 50-14 wc; (11) 35/14 mm; and (12) 50/10 dbc. (Acronyms used here: drb = dense roadbase; wc = wearing course; dbc = dense basecourse.)

Figures 1–5 show representative images from the first five of these classes.

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 a feature selection and multiple discriminant analysis approach, to support nearest neighbor (Mahalanobis metric) image querying.

1.3 The Vision Model Perspective

For robotic and industrial images, the objects to be detected and analyzed are usually solid bodies. An appropriate vision model for such images is therefore based on the detection of the surface edges (see e.g. [4]). 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 [20], 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, our vision model is as follows. For fine grained material, image characterization is based on image entropy, determined by resolution scale (thereby

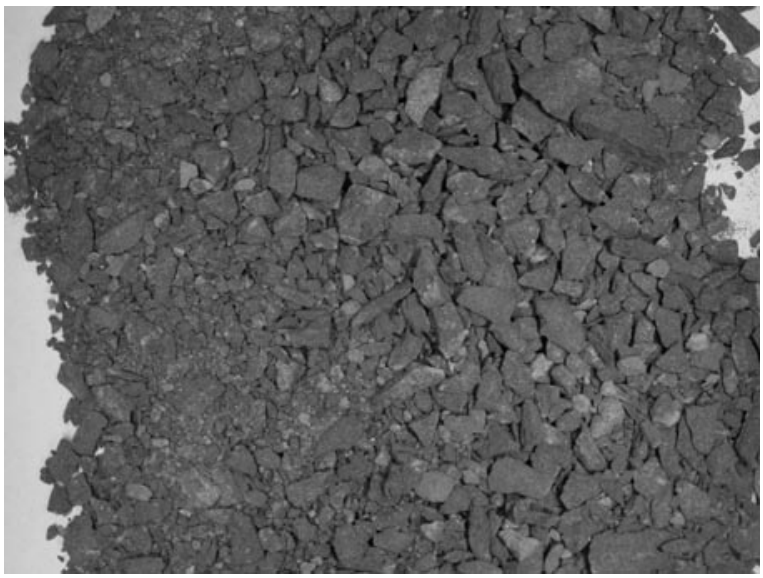

Fig. 1 Image from class 1. (We processed such an image as seen here, with non-influential "nuisance" regions near the image boundary.)

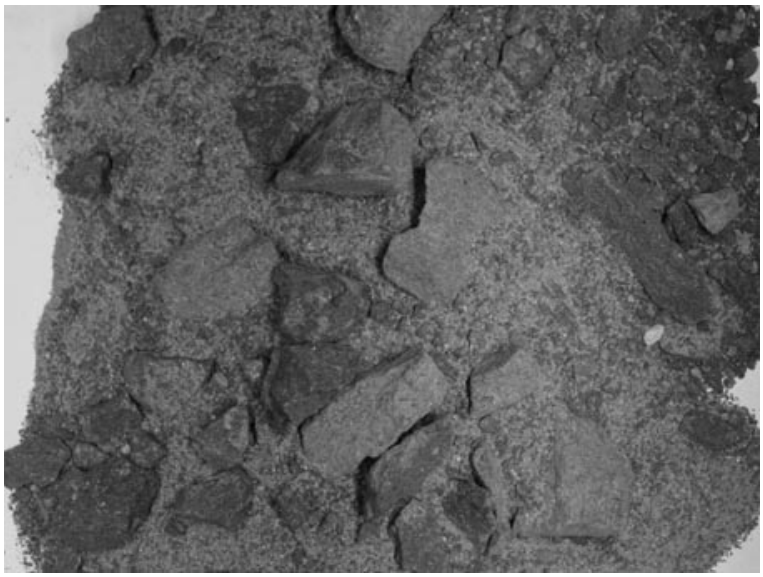

Fig. 2 Image from class 2.

catering for gradation in granularity size), and by spatial region (thereby catering for more coarse aggregate). Both spatial and resolution aspects are naturally implemented using a wavelet transform.

Furthermore given that smoothing takes place in the succession of band-pass filtered pixel data in the wavelet transform, we examine statistically significant wavelet coefficients at each resolution scale. Spatial (intra wavelet scale, and not inter scale) adjacency, only, is used to define (or more accurately express the contribution towards the values of) object features. If it were necessary to analyze such objects, individually rather than globally, then inter scale analysis would be carried out.

2 Feature Selection

2.1 Wavelet Transforms and Texture Features

A great deal of work has been carried out on automated segmentation of texture images since Cross and Jain [5] took the ideas of Besag [2] and applied them to realistic textures. Such work is based on a

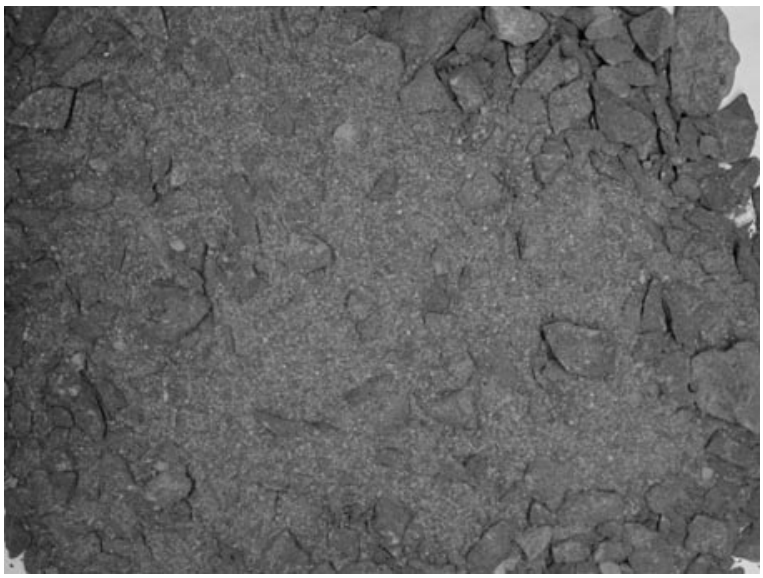

Fig. 3 Image from class 3.

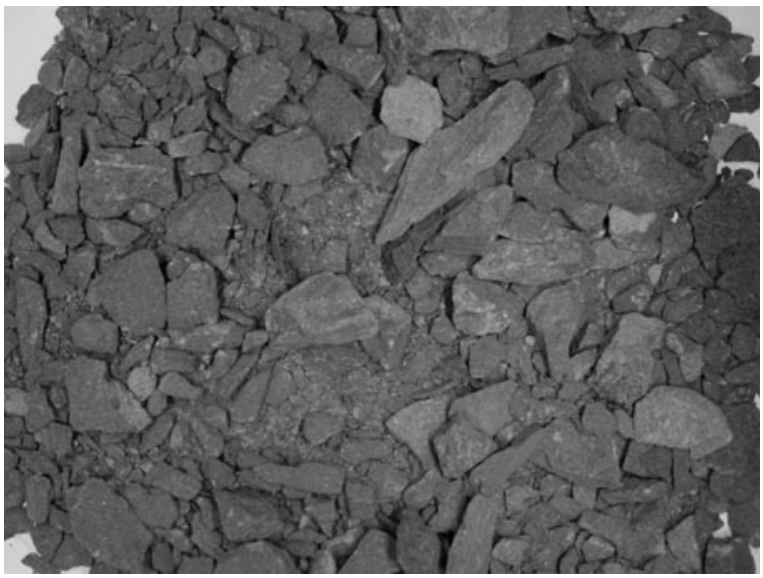

Fig. 4 Image from class 4.

Markov model of spatial context. The wavelet transform, being both a frequency and space transform, provides another way to model local spatial relationships. The wavelet transform has been often used to provide an embedding of the image in a multidimensional feature space. Furthermore, successive wavelet resolution scales express increasingly large neighborhood correlation. A clear advantage is offered by the wavelet transform (computation linear in the number of pixels, for a fixed number of resolution scales) over Markov modeling (based on iterative optimization). In previous work [12–14] we carried out image segmentation. Below in this section we will present our case for a feature-based characterization of images, with no need for segmentation.

In Unser [21], it is suggested that co-occurrence matrices and second order statistics may be best for segmentation of microtextures. The use of a wavelet transform for this purpose was first proposed by Mallat [10]. A redundant wavelet transform provides translation-invariance, and can therefore be of help in demarcating texture region boundaries. Unser [21] uses a wavelet frame filter bank, and in each band a sliding B-spline window which is convolved with wavelet coefficient energies. Therefore a windowed weighted average energy is used. The window size varies with resolution scale. KLT (Karhunen-Loève transform, or principal components analysis) is used for dimensionality-reduction. K-means is used for clustering. (We mention these because of our use of multiple discriminant analysis

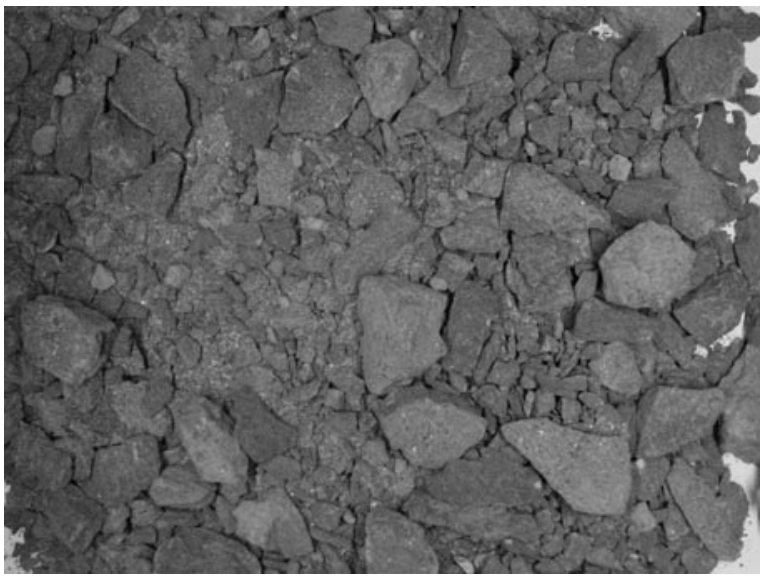

Fig. 5 Image from class 5.

as an alternative, integrating supervised classification, and dimensionality-reduction for visualization.) Markov spatial models are discounted by Unser [21] as being too time-consuming. Results on synthetic mixtures of Brodatz textures are discussed in relation to known "gold standard" labeling. A wide-ranging comparative study in [15] provides support for energy and for Gaussian-like properties of the "filter" function (even if clearly such support cannot be universally applicable). A close variant of energy is also used in [1].

Scheunders et al. [18] discuss multiband features, which they use with 3-band color data. In Livens et al. [9], energy in different bands is used. This does not provide image rotational invariance and for this a sum of energies over bands at a given resolution scale is proposed. Fatemi-Ghomi [6] uses window size related to resolution scale within which to define features, and she discusses adaptive window sizes.

Our requirements include the following.

1. Like previous work, we will define features based on second order moments. Further we will discuss the interesting special case where variance, energy and entropy are closely related: we show that entropy equals energy divided by variance in the special case of Gaussian-distributed pixel values.
2. We wish to avoid the use of windows, for which the size necessitates the specifying of an extra parameter.
3. We use a redundant wavelet transform because it avoids aliasing due to decimation, and, as noted above, a spatial model (correlation between neighboring pixels) is naturally part of it.

2.2 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 [6,10,9,21,18]. Our approach avoids any system parameter related to window size; and the undecimated wavelet transform used helps to avoid object aliasing.

We now briefly describe the features used: they will be discussed later in this section and in the following one.

- We use multiple scale image entropy to quantify aggregate granularity. Using 5 wavelet scales, from the B_3 spline à trous redundant wavelet transform [20,19], an entropy-per-scale was determined, and thus provided a 5-valued feature vector for each image. Among alternatives, properties of importance for us include rotational invariance (we are dealing with potentially very different types of texture compared to those considered by [15,1]) and a Gaussian-like scaling function.
- 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, and n 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, $j = 1, 2, \ldots, l-1$, have 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, at all events, $< \log_2 n$. The B_3 spline à trous redundant wavelet transform does not favor any orientation, and being redundant avoids decimation-related effects.

The feature set is defined from the resolution scale related decomposition as follows:

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

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

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

where σ_j is the noise (standard deviation) 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.

An initial approach by us to feature selection was based on an analysis of significant structures defined in the following way. If one assumes a simple statistical model, such as a Gaussian one, it is straightforward to determine the Gaussian parameters in the wavelet scales. Note that the wavelet planes or scales 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. While thresholding based on this noise model has good theoretical credentials, we use this framework here as a heuristic to distinguish between interesting and uninteresting signal. In our feature set of significant structures, 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.

Subsequently we bypassed this approach to use the entropy at each resolution scale, yielding 5 features per image. Note that the entropy necessarily uses a probability model for the wavelet coefficients, and on heuristic grounds we used a Gaussian model.

2.3 Larger Aggregate Characterization

We next propose an innovative approach to defining features based on coarse aggregate. Larger pieces of aggregate can be modeled using the "texture" oriented approach described above. We used, in addition to the features used so far, a set of features designed to characterize larger pieces of aggregate.

The additional features were those used previously, determined not from the given image but rather from the gradient image. The Canny edge detector [3] was used to produce the gradient image. Stronger gradients, of course, are yielded by outlines of coarse aggregate, which motivates our use of these features.

In all therefore, per image, we used 10 features.

Typically the edge finding, and entropy finding at the succession of resolution scales, for an image of the dimensions considered in our results, below, takes 8 seconds on a Sun Sparcstation 10. The entropy finding takes, typically, 6 seconds. Other operations considered below (assignment using the most important discriminant factors yielded by a multiple discriminant analysis) are effectively real-time.

2.4 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) [11,17]. 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. See also Hand's [7] case for "simple is best" in regard to choice of classifier.

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 centered 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.

3 Implementation and Evaluation

Figure 6 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 centering; 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 2.3 above.) For five successive classes of image, among the 250 images used, we had numbers of images misclassified as: 0, 0, 3, 7, 0.

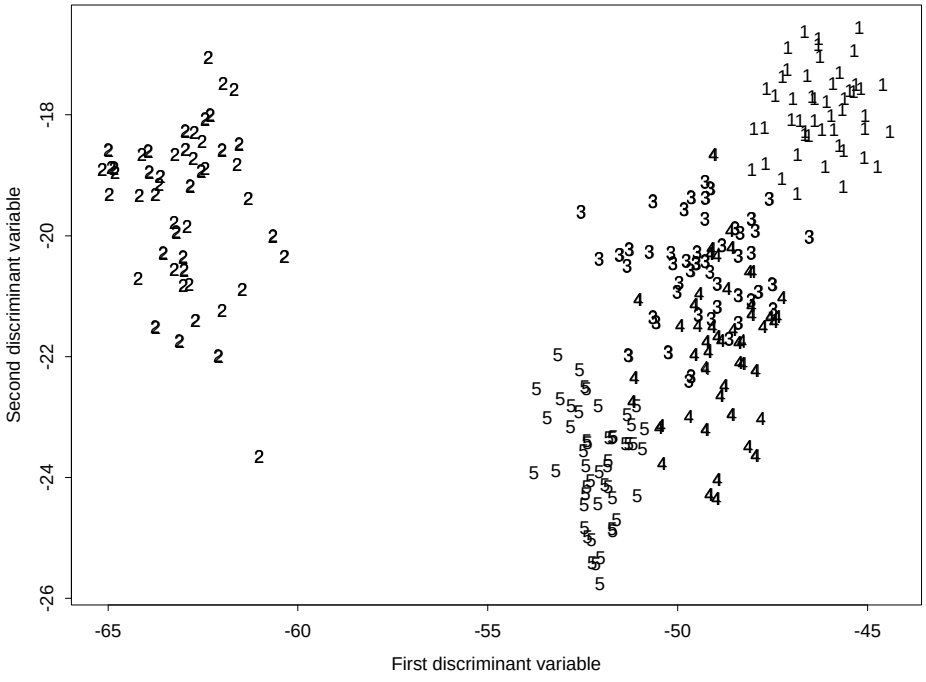

Fig. 6 Principal discriminant factor plane, with projections of images from classes 1 to 5.

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

1. We examined the denoising of the image data prior to analysis. This was not found to be of benefit.
2. Rather than the nearest mean classifier used in the multiple discriminant analysis, we also investigated nearest neighbor discrimination approaches (1-NN, 3-NN); and a multilayer perceptron. However the hyperlinear approach was found to give good results, and its operation was easily controlled and managed.
3. We worked on rebinned 454×340 images, obtained from the originally sized 2272×1704 images. We tested the processing of a battery of 600 images on the original 2272×1704 sized images. Days of compute time on a Sun Microsystems cluster gave us an approximate improvement of 2% misclassification. We saw no justification in continuing to process the original images in this way.
4. 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.

In order to study the needs for training and test set cardinalities, assessments were carried out and are displayed graphically in Figures 7 and 8. Class A (test) corresponds to 1 (training), B (test) corresponds to 2 (training), and so on. For Figure 7, 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 8, an average 95% success rate was obtained on the test sets.

Fig. 7 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. Feature space was 10-dimensional, defined from multiscale entropy.

Specifications of mixes were, in this work, within standard bands. We have now started to extend these results to ensure specification band coverage.

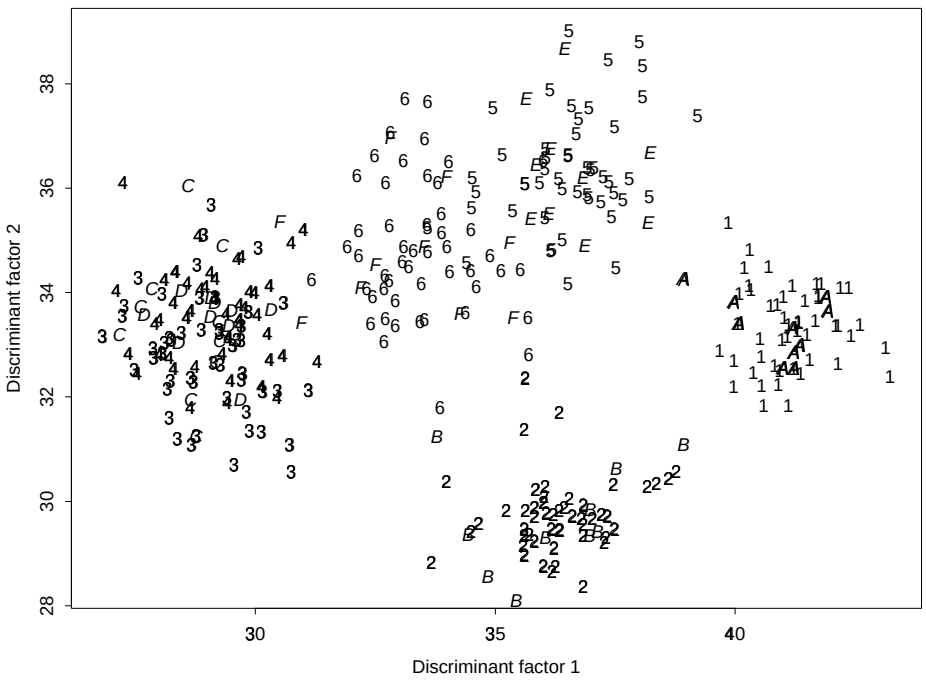

Fig. 8 Principal discriminant factor plane, with projections of images from groups 1 to 6. The training set made use of 50 images from each group. The test set made use of 10 images from each group. Feature space was 10-dimensional, defined from multiscale entropy.

3.1 General Context: Information Retrieval from Image Databases

While segmentation is an important objective (see our previous work in [13] or a general treatment in [1]) we have sought to avoid image segmentation in the work described in this paper, on the grounds of more directly, and computationally efficiently, addressing user needs.

Driven by generic multimedia database applications, there has been much work in recent years on the theme of content-based image retrieval. Support has included cultural heritage and museum applications, and personal and journalist digitized photograph collections. Querying in generic database applications has often involved used of color and object (sub-image) shape properties.

Retrieval from specialized databases (e.g. fingerprint databases, or astronomy image stores, or Earth observation databases) has usually availed of specialized image features. Users of information retrieval systems in these specialized areas are most often professional experts. Thus, for example, Earth observation imagery is usually queried on the basis of well-defined external properties: dates of observation, resolution scale of detector, geographic coordinates.

In the work described in this paper we are dealing with information retrieval support using a specialized image database of engineering materials. In this context, we have proposed and evaluated a range of general features, derived from the images. These features are defined from image resolution scale-based gradient, energy and entropy properties.

4 Conclusions

We have tested a new image content characterization approach, with excellent results on images of aggregates containing a wide range of object sizes and morphologies. Our algorithms, based on global, overall features, are computationally inexpensive, and scalable. We have shown in our experimental evaluation that these algorithms are robust and stable.

From the point of view of operational use in the difficult conditions of the construction industry, care and attention are required in the operational camera and lighting environment, and in limiting vibration effects, so that a comprehensive field of view is obtained.

References

1. M. Acharyya, R.K. De, and M.K. Kundu. Extraction of features using M-band wavelet packet frame and their neuro-fuzzy evaluation for multitexture segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25:1639– 1644, 2003.
2. J. Besag. Statistical analysis of dirty pictures. *Journal of the Royal Statistical Society, Series B*, 48:259–302, 1986.
3. J. Canny. Computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8:679–698, 1986.
4. H. Choi and R.G. Baraniuk. Multiscale image segmentation using wavelet-domain hidden Markov models. *IEEE Transactions on Image Processing*, 10:1309–1321, 2001.
5. G. Cross and A. Jain. Markov random field texture models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 5:25–39, 1983.
6. N. Fatemi-Ghomi. *Performance Measures for Wavelet-Based Segmentation Algorithms*. PhD thesis, Surrey University, 1997.
7. D.J. Hand. Academic obsessions and classification realities: ignoring practicalities in supervised classification. In *Classification, Clustering, and Data Mining Applications*, pages 209–232. Springer, 2004.
8. P. Hobeda. Krossningens betydelse på stenkvalitet, starskilt med avseende på kornform. Literaturstudie Nr. 050001, Statnes vag-och Trafikinstitut, VTI, Linköping, Sweden, 1988.
9. 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.
10. S.G. Mallat. A theory of multiresolution signal decomposition: the wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11:674–693, 1989.
11. F. Murtagh and A Heck. *Multivariate Data Analysis*. Kluwer, 1987.
12. F. Murtagh, Xiaoyu Qiao, D. Crookes, P. Walsh, P.A.M. Basheer, and A. Long. Benchmarking segmentation results using a Markov model and a Bayes information criterion. In *Opto-Ireland 2002: Optical Metrology, Imaging, and Machine Vision, Proc. SPIE, Vol. 4877*, pages 248–254, 2003.
13. X. Qiao, F. Murtagh, D. Crookes, P. Walsh, P.A.M. Basheer, and A. Long. Machine vision methods for the grading of crushed aggregate. In *Opto-Ireland 2002: Optical Metrology, Imaging, and Machine Vision, Proc. SPIE, Vol. 4877*, pages 264–270, 2003.
14. X. Qiao, F. Murtagh, P. Walsh, P.A.M. Basheer, D. Crookes, and A. Long. Image processing of coarse and fine aggregate and SEM cross-section images. In *Proceedings of the International Workshop on Structural Image Analysis in Investigation of Concrete, SIAIC'02*, pages 231–238, 2002. Warsaw, Poland.
15. T. Randen and J.H. Husoy. Filtering for texture classification: a comparative study. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21:289–290, 1999.
16. V. Reinhardt. Schlagfester Splitt 8–11mm oder stabiler Asphaltbeton 0–12mm. *Bitumen, Teere, Asphalte, Peche und verwandte Stoffe*, 1969. Nr. 11.
17. J.M. Romeder. *Méthodes et Programmes d'Analyse Discriminante*. Dunod, 1973.
18. 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.
19. J.L. Starck and F. Murtagh. *Astronomical Image and Data Analysis*. Springer, 2002.
20. J.L. Starck, F. Murtagh, and A. Bijaoui. *Image Processing and Data Analysis: The Multiscale Approach*. Cambridge University Press, 1998.
21. M. Unser. Texture classification and segmentation using wavelet frames. *IEEE Transactions on Image Processing*, 4:1549–1560, 1995.
22. Weixing Wang. Image analysis of aggregates. *Computers and Geosciences*, 25:71–81, 1999.
23. W.X. Wang and O. Stephansson. Comparison between sieving and image analysis of aggregates. In *Measurement of Blast Fragmentation*, pages 141–149. Balkema, Rotterdam, 1996.